



C o m m u n i t y E x p e r i e n c e D i s t i l l e d

Game Data Analysis – Tools and Methods

A data-driven approach to video game production

Coupart Thibault

www.allitebooks.com

[PACKT]
PUBLISHING

Game Data Analysis – Tools and Methods

A data-driven approach to video game production

Coupart Thibault

[PACKT]
PUBLISHING
BIRMINGHAM - MUMBAI

Game Data Analysis – Tools and Methods

Copyright © 2013 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing, and its dealers and distributors will be held liable for any damages caused or alleged to be caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

First published: November 2013

Production Reference: 1181113

Published by Packt Publishing Pvt.Ltd.
Livery Place
35 Livery Street
Birmingham B3 2PB, UK.

ISBN 978-1-84969-790-3

www.packtpub.com

Credits

Author

Coupart Thibault

Project Coordinator

Ankita Goenka

Reviewers

Kevin Corti

Kevin Flood

Bruno Santos

Proofreader

Joel Johnson

Indexer

Mariammal Chettiyar

Acquisition Editor

Martin Bell

James Jones

Production Coordinator

Kirtee Shingan

Commissioning Editor

Neil Alexander

Cover Work

Kirtee Shingan

Technical Editors

Novina Kewalramani

Anita Nayak

Harshad Vairat

Cover Image

Disha Haria

Copy Editor

Roshni Banerjee

Janbal Dharmaraj

Aditya Nair

Deepa Nair

Karuna Narayanan

Alfida Paiva

About the Authors

Coupart Thibault had been studying town planning and statistics for years before he developed an in in-game data analysis. Switching from the real world to the virtual world, he studied game design and completed his transformation with his first professional experience at Corexpert. Working on a database of more than 20,000 players, he has experience with the study and analysis of data in order to deliver various insights for the company.

I would like to express my very great appreciation to Mr. Olivier Masclef for the time spent at the Gamagora video game school, my supervisor Mr Jérôme Vuillemot at Corexpert, and Mr. Johan André-Jeanville (from a consulting firm for LudoStat data analysis) for his advice on various topics related to game data analysis.

About the Reviewers

Kevin Corti is a games and digital media professional who has experienced the designing and delivery of well over 100 mobile, social, and serious games, and e-learning and web application projects since 1997. Kevin has founded four digital businesses including SoshiGames, a UK-based social music games publisher, and PIXELearning, one of the world's first dedicated serious games developers. He was previously Head of Studio at DNA Interactive and is now Chief Evil Officer at Evil27 Games, and particularly focuses on social mobile games discovery, retention, and monetization.

Kevin Flood has worked in the console game, social gaming, and Internet gambling sectors as a CEO, CTO, and Managing Director. He currently operates a consulting company, building social games and advising social game developers on ways to improve revenue per game and player. He also works with land-based casino operators on ways to launch social and real money casino platforms.

Kevin writes a gaming blog kevinflood.blogspot.com that covers topics including social gaming, social casino, and "real" money gambling.

Kevin speaks at gaming conferences worldwide addressing the growing importance of games as a form of entertainment that continues to grow, evolve, converge, and generate a substantial amount of revenue.

Bruno Santos main background is Computer Science, which has lead him through a more technical pathway in game design since his early years in college. Nowadays, he runs a two-man game studio based in Portugal.

www.PacktPub.com

Support files, eBooks, discount offers and more

You might want to visit www.PacktPub.com for support files and downloads related to your book.

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.PacktPub.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at service@packtpub.com for more details.

At www.PacktPub.com, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.



<http://PacktLib.PacktPub.com>

Do you need instant solutions to your IT questions? PacktLib is Packt's online digital book library. Here, you can access, read and search across Packt's entire library of books.

Why Subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print and bookmark content
- On demand and accessible via web browser

Free Access for Packt account holders

If you have an account with Packt at www.PacktPub.com, you can use this to access PacktLib today and view nine entirely free books. Simply use your login credentials for immediate access.

Table of Contents

Preface	1
Chapter 1: Context and Themes in Games	5
The rise of game analytics	6
Themes in games	7
The desire for reward	8
Ownership	8
Reputation	9
Achievement and collection	10
The desire for challenge	10
Complexity and difficulty	11
Competition between players	11
Desire for imagination	12
Discovery	12
Emotions and sensations	13
Immersion, story, and universe	14
Desire for entertainment	14
Distraction	14
Romping	15
From themes to engagement	15
Video game as a service	15
Free-to-play and engagement	16
References	17
Summary	17
Chapter 2: Common Key Performance Indicators	19
Definition and framework of Key Performance Indicators	19
Criteria	20
Structure	20
Acquisition of new players	21
Retention of players	23
Monetization of players	24

Working regularly with KPIs	25
Summary	26
Chapter 3: Environment and Tools for Data Analysis	27
Typical programming environment for data mining and storage	27
MySQL	27
NoSQL	28
Hadoop and Hive	29
Tools available on the market for quick data mining	29
Available free tools	30
Facebook Insights	30
Google Analytics	31
Commercial solutions	32
Kontagent	32
Honey Tracks	32
Flurry Analytics	33
Tools available for analysis	33
Open source tools	33
R-Project	34
Summary	36
Chapter 4: Game Analytics and Generation of Content	37
The A/B testing	37
The multivariate testing	38
Recommendations for good practices	38
Case study – monetization pop up	40
Inventory of each feature	40
Concrete examples of versioning	42
First example	42
Second example	44
Summary	44
Chapter 5: Advanced Analysis and Statistical Methods	45
General statistical description	45
Central tendencies	46
Dispersion tendencies	47
Statistical distribution and laws	48
Correlation and regression between variables	49
Types of variables	49
Chi-squared test	50
Linear regression	51
Logistic regression	52
Machine learning	52
Definition	53

Supervised learning	53
Unsupervised learning	54
Summary	55
Chapter 6: Data Visualization	57
Recommendations for good practices	57
Basic recommendations	58
Typical data visualization tools	58
Line chart	58
Bar chart	59
Round chart	59
Heatmap	60
Graphic semiology	60
Typical traps of data visualization	61
The choice of the scale – values on axis	62
The choice of the scale – equivalence and units between variables	63
Summary	64
Chapter 7: Limits of Game Data Analysis	65
Which game analytics should be used	65
Game analytics as a tool	65
Game analytics must serve your team	66
When starting, focus your attention on simple practices	66
What game analytics should not be used for	67
Keep away from numbers	67
Practices that need to be avoided	67
Summary	68
Index	69

Preface

Game Analytics is a general guide about the tools and the methods used in the data analysis applied to the video game field. In the seven chapters of this book, we will see a wide range of topics included in the domain, such as how to use game analytics to improve features and game design, balancing, technical tools, reporting of Key Performance Indicators, and some advanced statistical process.

What this book covers

Chapter 1, Context and Themes in Games, contextualizes the subject inside the current state of the video game industry. We will outline some of the economic trends that affect it, and explain why game analytics is going to get bigger and bigger. Moreover, we will highlight the topic under the angle of the sociological approach, providing a list of important gaming appetencies already identified by previous researches.

Chapter 2, Common Key Performance Indicators, gives a complete list of the most useful indicators, which are currently used in the video game industry, classified by type, from the acquisition and the retention of players to its monetization.

Chapter 3, Environment and Tools for Data Analysis, highlights tools dedicated to data mining and the treatment of data. We will have a few words about advantages of relational and non-relational databases, and explore some commercial solutions as well as free solutions.

Chapter 4, Game Analytics and Generation of Content, describes in detail one of the most famous process initiated with the rising of analytics, the A/B testing, and the generation of content based on data analysis. Some very concrete examples will be highlighted in which area of your game the A/B testing is usually the most interesting to perform.

Chapter 5, Advanced Analysis and Statistical Methods, tries to give an overview of different tools used in the statistical field and how they can be used for video game analysis. We will see the most common indicators, such as the mean, median, and standard deviation, and we will also explore some processes from the machine learning field, such as clustering algorithms.

Chapter 6, Data Visualization, gives some of the best practices about visualization of the data and some of the traps you need to avoid with it. We will talk about graphic semiology and present some concrete example problems, such as the choice of scale for a chart.

Chapter 7, Limits of Game Data Analysis, tries to give a final word on the topic and will highlight some of its limits inside the framework of the video game industry.

What you need for this book

You do not need anything in particular for this book. If you want to remake and test some of the statistical examples inside the book by yourself, a basic knowledge of Excel and the software installed on your computer will be useful.

Who this book is for


This book is aimed at any person working in the video game industry who wants to improve their knowledge on the new topic of game analytics. You do not need any particular advanced skills in data mining or data analysis to read this book, just a familiarity with statistics might be a plus, especially for *Chapter 6, Data Visualization*. Advanced data analysts should also find some interesting pieces of reflection through the game-design approach (*Chapter 2, Common Key Performance Indicators*, and *Chapter 5, Advanced Analysis and Statistical Methods*), whereas product managers will be glad to have an overview of the tools (free and commercial) currently available.

Conventions

In this book, you will find a number of styles of text that distinguish between different kinds of information. Here are some examples of these styles, and an explanation of their meaning.

New terms and **important words** are shown in bold. Words that you see on the screen, in menus or dialog boxes for example, appear in the text like this: "The objective, of course, is to have the highest number of clicks on the **Buy** button".

 Warnings or important notes appear in a box like this.

 Tips and tricks appear like this.

Reader feedback

Feedback from our readers is always welcome. Let us know what you think about this book – what you liked or may have disliked. Reader feedback is important for us to develop titles that you really get the most out of.

To send us general feedback, simply send an e-mail to feedback@packtpub.com, and mention the book title via the subject of your message.

If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, see our author guide on www.packtpub.com/authors.

Customer support

Now that you are the proud owner of a Packt book, we have a number of things to help you to get the most from your purchase.

Errata

Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you find a mistake in one of our books – maybe a mistake in the text or the code – we would be grateful if you would report this to us. By doing so, you can save other readers from frustration and help us improve subsequent versions of this book. If you find any errata, please report them by visiting <http://www.packtpub.com/submit-errata>, selecting your book, clicking on the **erratasubmissionform** link, and entering the details of your errata. Once your errata are verified, your submission will be accepted and the errata will be uploaded on our website, or added to any list of existing errata, under the Errata section of that title. Any existing errata can be viewed by selecting your title from <http://www.packtpub.com/support>.

Piracy

Piracy of copyright material on the Internet is an ongoing problem across all media. At Packt, we take the protection of our copyright and licenses very seriously. If you come across any illegal copies of our works, in any form, on the Internet, please provide us with the location address or website name immediately so that we can pursue a remedy.

Please contact us at copyright@packtpub.com with a link to the suspected pirated material.

We appreciate your help in protecting our authors, and our ability to bring you valuable content.

Questions

You can contact us at questions@packtpub.com if you are having a problem with any aspect of the book, and we will do our best to address it.

1

Context and Themes in Games

For several years now, the video game industry has experienced significant changes in its mode of distribution and content. The most important part of these changes comes from the digitization of the product, coupled with the democratization of the Internet, whereas the medium was essentially dedicated to the television through the golden age of game consoles (1990 to 2005). We can now play video games on multiple platforms, which still includes consoles and computers, but also includes mobiles and tablets.

Moreover, the acquisition of a game does not require us to buy a box from a store anymore. Instead, we can download it from the cloud. Thanks to the platforms, such as the Apple Store or Steam, we can directly play games on our Internet browsers without having to wait.

The rise of social networks, such as Facebook, has also altered the way we play games today, changing the type of content delivered to the audience exactly as mobiles have done. They directly integrate an application into the structure of the social network, and these social games may be played asynchronously and strongly promote social interactions between players.

On the other hand, mobile games are supposed to propose a short temporality where the user will have action, fun, and rewards after a few minutes; you have it in your pocket and you will play games while waiting for the bus, or anything similar.

Following the evolution of platforms, game design practices have also seen a renewal. Most of those new games rely on a business model called **free-to-play**. This expression means that the game is available for free and that the player will have the possibility to buy virtual goods in the game with real money. Thereby, the entire experience needs to take this virtual market in account and the implications in terms of game design; what kinds of goods are purchasable, which content might be unlocked, and what are the best prices for an item for the user as well as for the general profit of the product.

To summarize this idea, this new model has brought new challenges for the developers because the objectives today are no longer to launch the best AAA product before Christmas, but to follow and constantly improve a game whose benefit does not rely on its entry price anymore.

These evolutions had and have huge impacts on several aspects of the media and are definitely breaking the current balance of the video game industry, with new surprising successes every month and older publishers hurrying to adapt to the new digital economy. According to a study published in May 2012 from the Japanese magazine publisher *Enterbrain*, digital sales (including online and downloadable games) now represent 58 percent of the global video game turnover, slightly more than its retail sales, for a total amount of approximately 30 billion euros.

The rise of game analytics

Among all impacts, one field is particularly underrated and still growing; the use of game analytics in the creation and update process of video games. As stated previously, having a game which is always connected to the Internet allows you to make frequent updates or patches on it, such as changing some old features and adding new content, and moreover collecting data permanently about user behavior.

Far from trivial, these new tools are progressively changing the process of creating video games in depth. Back in the old days, a video game was a finished product when it went live and it has passed many usability tests and player tests, as well as technical tests (bugs identification, and so on). This statement is especially true about the old console business model and the back and forth between a video game studio and the owner of the hardware (Nintendo, Sony, or Microsoft), and a game had to respect a certain number of quality requirements before it became marketable.

However, many video games, especially in the social and mobile sphere, now go live, unfinished on purpose and are improved through user data feedbacks, weeks after weeks, months after months.

We can definitely attribute the starting point of this method with the rise of Zynga and its **Farmville** game, published on Facebook in 2009. As *Brian Reynolds*, former chief game designer, stated on *VentureBeat* on February 2013:

"The prototype-and-iterate method of game design I'd learned (...) seemed like it might fit with the completely new watch-metrics-and-tune methods that Zynga had been perfecting (.....). And the challenge of creating game experiences with "perfectly rounded edges" – smooth and friendly for the masses – has been one of the most interesting challenges of my career."

– *Brian Reynolds: On Zynga, games, and the future*

Brian Reynolds has been the lead game designer of Zynga for years and this quote really shows how the new business models have broken the rules; neither the approach nor the design processes have been spared by the winds of change and game analytics is currently seeing an increasing popularity, which surpass the field of social gaming and conquer more and more companies every month.

Following this idea, talking about game analytics through the video game field does not mean focusing only on the tools themselves, but it also means considering these tools in a wider strategy. A strategy includes the framework of a studio, for example, which joins with the different professions that composed it and its objectives; for this, being data driven definitely requires investments and efforts.

Basically, the purpose of game analytics is to solve a very old question that has existed since the beginning of the media: why do people play a game and what do they enjoy the most while playing it?

Beside being a subjective intuition, game analytics definitely appeared to be the best way to answer this question. Considering this field covers the process of collecting and analyzing data (especially in-game data) in order to understand user behavior and support decisions about the development of a game, it has its foundation in the field of statistics, which existed before there were video games. For this reason, it already offers a very large range of methods and tools.

The following statement is an advantage and a disadvantage at the same time:

- Many things have already been experienced, but on the other hand, we need to deal with the ocean of terms and methods that exist, and choose the best tool according to its own needs

This is the reason for this book; to propose a list of tools and methods from the data analysis field which fits the video game problem as much as possible and illustrate it with concrete examples from different products, especially from social and mobile.

The next section will be focused on user gaming appetencies, definitely a social and psychological point of view, and will be an excellent starting point before describing the tools and methods of game analytics.

Themes in games

Following this contextualization, we will now evaluate a typology of several themes, which are important to know before starting an analysis on video games. These themes are inspired at various levels by a part of the work made by psychologists, business consultants, and game designers over the last ten years. I would especially like to quote from *Nicole Lazzaro (2004), Four keys to more emotion without story*, the **DGD1** model according to *Chris Bateman (2005)*, and the Gamasutra article *Five ways games appeal to players* from *Jason Tocci (2012)*.

Other references related to this topic may be found by the reader at the end of this chapter, if he wants to go further with them.

Many instances of concrete research, like those stated previously, have already been done on this topic, and we will try to innovate in the following two ways:

- Usually, most of those studies lack one thing; rich and concrete situational examples from video games. Here we will try to avoid this default case and instead propose many video game examples.
- Moreover, most of the chosen examples in this chapter come from the mobile and social sphere, and it will be a great occasion for the reader to discover how old research on traditional console games can be associated and highlighted with the new digitized video game industry.

This chapter should be useful for the following people:

- Young data analysts starting a career in the video game industry who want to improve their affinity with the media; analyzing numbers is one thing, but having a good knowledge of mechanics involved by game design will definitely help to start. For the same reason, this chapter should be interesting for data analysts who apply for a job and want to avoid some typical traps asked by recruiters.
- From another perspective, scientists and academics will find many video game references in this chapter that can help them to find suitable examples for their argumentation.

The desire for reward

For many players, the reward after completing a task is the most important thing in a game. It drives the way they play the game and while trying to achieve an objective, they will have the reward in mind. This reward can have different signification, and we can point out three subcategories or keywords: ownership, reputation, and achievement.

Ownership

The desire for ownership relies on the item that has an intrinsic value in the game and for the player. It will mostly be an item for his avatar, but considering city management games such as Farmville (Zynga, 2009), it could be a building or a decoration.

Many games rely their general experience only on this appetency that is, cash machine games (virtual or not), and in a more subtle way sandbox games such as Minecraft, where the player has total freedom and numerous tools to create and build their own valuable reward.

In the free-to-play game sphere, the way this desire is introduced into the game is a very important factor for its success; the largest income for them comes from the sale of virtual goods to the player. Moreover, this desire is used in social and mobile gaming in many situations in order to increase revenue as well as to retain players, and we are going to see the following few examples:

- In many Facebook games, the player gets a special reward for connecting himself to the game regularly. Usually, the temporality for this reward is on a daily basis, and the reward desire is used to improve gaming frequency.
- Some games even go further; on Galaxy Online 2 (publisher I Got Game), a player will receive a reward every 15 minutes he stays connected to the game, this is to encourage people to keep the game running on browser even if they do something else.
- Another very typical mechanism in the social gaming is to reward the player for inviting one of his (Facebook) friends to the game. The reward desire here is used to improve the acquisition process of new players.

Reputation

This subcategory is focused on the social recognition obtained; thanks to successful actions inside the game. After completing a task, some players enjoy sharing their success in front of their social network, and this desire drives the way they play games. This idea has been definitely highlighted by the success of Facebook as a gaming platform, because many game mechanics from Facebook games rely on the facility of social interactions between players.

One of the most typical Facebook game features is to propose to a player to display his success or results in the game on his Facebook wall. This will alert his Facebook friends and may encourage them to play the game as well.

Many "reputation desire" mechanics can be found also in **Massive Multiplayer Online Role-Playing Game (MMORPG)**, for example, the equipment of an avatar shows the success and the progression of the player in the game.



Some MMORPGs go further, in World of Warcraft (Blizzard Entertainment) players have the possibility to complete special challenges in order to get a special title beside their avatar name. These titles do not really have any interests for the core progression but they show to the community that "they have completed this special challenge", thereby they constitute a source of celebrity status for the player.

Achievement and collection

The achievement subcategory refers to the pleasure felt when a task is completed and when nothing is left behind. This is very close to the collection pleasure, which is about collecting every valuable piece of something; it can be stamps, paintings, jewelry, and so on, and from a video game's point of view, it could be cards, items, achievements, and Pokemons.

Usually, collection and achievement mechanics are quite useful in order to add some long term objectives or additional content.



Many adventure games add hidden collectible bonuses among different levels of the game, with a special achievement for the player who collects them all.

Looking at social and mobile gaming, these mechanics are quite popular too and you can find special achievement bonuses in games, such as Angry Birds (having all of the five stars on each level), Rayman Jungle Run on IOS (Ubisoft).

Some games go further with this principle of collection and use it to add some fun inside their in-game items economy. In Warframe (Digital Extremes), a free-to-play third-person shooter, skills and abilities take the form of cards, which drop randomly on monsters, and they can be combined and merged in order to get even more valuable cards.

The desire for challenge

Before earning rewards, a player needs to complete a task; finishing a level and completing a mission are some of the most obvious leitmotifs in video games, but more precisely these objectives encourage the player to learn and assimilate the rules and mechanics of gameplay, and thus players are challenged by the game.

Following this idea, while the previous category was focused on the type of reward earned, this category is focused on the task itself and is directly involved in an important game-design theme, which is the flow channel. Apart from all psychological considerations, flow channel is quite simple to understand. If the game is too difficult, the player will be discouraged. On the other hand, if the game is too easy, he will be bored, and in order to keep the player involved in your game, you need to have a progression curve that fits his sensibility as well as possible.

Right now, this category is about the desire of challenge from players. What is important to notice is that some players really enjoy being challenged, and as we will see we can highlight two kinds of challenges: a challenge from the game and challenges between players (competition).

Let's consider these two subcategories in detail.

Complexity and difficulty

The promise for intellectual challenges, hard tasks, and big challenge inside the game may provoke many sensations for a player, including the production of endorphin and adrenalin, for example. Considering previous work on the subject, these emotions were easily attributed to "hardcore gamers" looking for challenge and intense action, but if we expand our view, we can notice that mobile and social gaming also offer some interesting examples close to this theme.



Brain Buddies (2009), the first game of the Wooga franchise, was composed of a series of mini-games where the final score was supposed to represent the level of your Intellectual Quotient. This was definitely a good reason to introduce some difficulty in the game without frustrating the players, because the initial game promise is an intellectual challenge. Combined with the typical sharing score option on Facebook, the success was quite immediate.

Overall, the way a challenge and progress curve is introduced inside a free-to-play game is determinant because many business models include some items that can bypass or diminish the difficulty of the game, so the equation is definitely tough and not too easy because players will be bored; it should be hard enough to encourage a part of the player database to buy virtual goods to help them progress in the game, but not too hard because players might be discouraged and will stop playing. As we will see later in the book, game metrics is definitely the best way to find the best balance for that kind of problem.

Competition between players

This subcategory is related to the pleasure of competition and how some players love to measure their strength against each other. Like the previous subcategory, this theme used to be easily attributed to hardcore gamers, but the rise of mechanics, such as sharing scores on Facebook, shows that casual games also have a real interest in developing (asynchronous or not) competition inside their community.



Diamond Dash, from the game publisher Wooga, motivates its community by updating a weekly tournament, where the score of a player is displayed next to his Facebook friends' scores. At the end of the week, the player with the best score receives a special reward, and a new tournament starts. The game is on Facebook as well as on mobile and scores from both platforms are harmonized.

On the other hand, games for hardcore gamers have also overcome a major step with the success of League of Legends (Riot Games), which has popularized the e-sport aspect of its game; it currently has high value tournament prizes, real time reporting during matches, and strong structure for streaming.


Desire for imagination

This category relies on the power of video games to create a universal story and immersion into the game as well as gameplay sensations. Usually dedicated to AAA games with 3D universe and advanced visual effects, we will see that this theme is also present and relevant inside the social and mobile gaming sphere.

We can highlight three subcategories here, namely discovery, universe, and emotion.

Discovery

Next to traditional console games, some very simple games from the social sphere also have their part of discovery for the player. Something very common and very simple is to propose different graphic background to the player while he progresses through the levels of the game. It might sound negligible, but a simple graphic background brings an additional reward to the player for completing a level; some players definitely enjoy this kind of attention and their interest in the game increases.

 Zynga Slingo (Zynga), Bubble Witch Saga (King.com), and many other games, which have a level-progression structure, proposed a large diversity of graphic background depending on the level the player is, even if gameplay mechanics do not see any major evolutions from a level to another.

The following screenshots shows two different graphic level backgrounds from Zynga Slingo, and as we can see, the gameplay framework remains the same but the atmosphere is clearly different:



Emotions and sensations

Emotion in games is a very controversial subject, and here we will consider that the word emotion refers to immediate feelings provoked by the in-game actions, a sort of immediate fun. Usually, these events imply the presence of surprise and encourage funny interpretations of game mechanics.



Many successful dexterity games on Facebook or mobile use visual effects to add special events for the player according to his actions. For example, being quick enough at destroying diamonds in Diamond Dash (Wooga) may provoke a flame event for a few seconds where points earned are doubled and the whole game board inflames itself, increasing the adrenalin for the player.

The following screenshot shows the flame event on Diamond Dash (Wooga):



Immersion, story, and universe

Similar to cinema, this subcategory relates to the narrative quality of a game and to the charisma of its characters. Even if video games have the advantage of interactivity, having a good storyline and quality characters are very helpful in increasing the player's involvement in the game. Considering mobile and social gaming, the relevance of this theme will depend on the choice of game design adopted for the game. Some games do not really need a story; others will build the progression of the player through it.



Hidden object games, such as Pearl's Peril (Wooga), or investigation games, such as Criminal Case (Pretty simple), are focused on the story to encourage the player to progress further.

Besides the story, characters in social games are sometimes the subject of the A/B testing, which character fits the audience of the game the best? A typical case is to propose different archetypes during the tutorial phase (usually a character helps the player to understand basics of the game) and to look at data to identify which one is the most successful.

Desire for entertainment

Very simply, this category is reliable for the stabilization of mind and the distraction that comes from the act of playing. Inherent to any kind of games, mobile is certainly the best platform for this theme because of its usage. People play games on mobile while waiting for a bus, at a metro, during a meeting, and so on. However, previous examples were mainly from Facebook, but we will use examples from mobile gaming for this category.

Distraction

Video games are sometimes a very simple and efficient way to spend time, especially when it is idle time. Moreover, it is also a way to avoid thinking about unpleasant things, and it offers a step away from reality. Putting this pleasure as a determinant factor for a mobile game implies many things in terms of game design. A game may last for a few minutes and the progress curve temporality must fit the usage of mobile.



Monster World (Wooga) is a management-economic game like Farmville, which has two different versions, one on Facebook and one on mobile, and both are not harmonized, unlike Diamond Dash. The explanation is simple; mobile and Facebook usage temporalities are not the same, and growth time for plants as well as to collect rewards must be different on both platforms. This statement does not hold true for Diamond dash, where one game lasts for one minute on both platforms.

Romping

Games, sometimes, are a good way to generate an overflow of energy and pour out negative feelings and destructive desire. Many products use the mechanics of bringing games and toys closer to increase a player's affection. Mobile games offer us the best examples for this mechanics here as well.



Fruit Ninja (Halfbrick Studio) and Angry Birds (Rovio) are certainly the most famous examples where players will have an obvious pleasure at destroying things. In Fruit Ninja, you will cut different kinds of fruits in pieces, whereas you will throw funny little birds on bad pigs in Angry Birds.

From themes to engagement

Themes that we have seen in the previous section are very important in a free-to-play environment. We must now describe this new rising model and illustrate why this kind of game design knowledge can help in many concrete cases.

Video game as a service

Thanks to the arrival of new technologies, such as the cloud gaming, social network, and mobile platform, the ways to provide entertainment are now very large and diffuse. In the current state of the industry, video game should now be considered as a service and not as a finished product anymore. Through regular updates and the use of data, an average game can become good and meet a certain amount of success. Thus, the player will be profitable, not just after it has been published. And as you want your service to be the best for your audience in order to increase your income, you would want to identify the category of themes your game belongs to. Ask the following questions to yourself:

- Which theme does your game design contain?
- Do players prefer the challenge over the exploration in your game? (this question needs to be answered according to your themes)
- Considering your data and the answers to the previous questions, do you think the service you deliver to your player base is relevant?

Following this reasoning, we will review in detail why this question of themes becomes even more important in a free-to-play context.

Free-to-play and engagement

In a free-to-play game, the main source of income is provided by the sale of virtual items to the players. While other sources of income, such as advertising, should not be forgotten, this one remains the most important. Talking about the notion of service again, virtual items that you are selling in the game are supposed to increase the quality of the service. They can provide help, facilities, or additional content to their buyer. This is where the notion of gaming appetencies becomes interesting again, because a virtual item can definitely improve a particular theme of your product.

As a studio, if you have online games based on the free-to-play model, ask yourself the following questions:

- Do purchasable items present in my game facilitate every kind of theme improvements? For example, may a player have more content to explore if he chooses to spend money? May a player have more chances to find treasures if he chooses to spend money? The list goes on according to the themes we have studied previously.
- According to the data, which sort of virtual item is the most successful? Is the answer what you expected it to be according to your original game design? If not, what could be the reason for that?

In order to help you answering the preceding questions, the following is a typology of saleable items that fit the gaming themes we have studied before:

- **Desire for ownership:** Game items available for purchase with real money boosts the resources, increasing the player's ability to buy new game items.
- **Desire for reputation:** Social sharing and additional social features could be unlocked with real money. Note that this sort of feature is usually free in order to improve the vitality of a game.
- **Desire for collection:** If your game proposes collectable items, you can sell items that will increase the chances of finding rare items. Note that this might be better than just selling the rare items because it adds the pleasure of the random factor and preserves the value of the items in the eyes of players.
- **Desire for complexity and difficulty:** A hard mode of your game dedicated to expert players could be unlocked with money.
- **Desire for competition:** Player versus player mode could be unlocked with real money. Note that, as for the social features, letting this mode for free can help your game conquer its audience.
- **Desire for discovery:** We can link this theme to any purchasable items that allows more content to be played. Additional levels are certainly the best example.
- **Desire for emotions:** Special visual effects, such as a shader that makes the screen shake when the player is hit or additional particles effects, could be

- purchasable with money. It can also be a new look for the avatar of the player.
- **Desire for immersion:** Additional content related to the background and the storyline of the game could be purchasable with money.
 - **Desire for distraction:** Purchasable items tend to decrease the difficulty of the game; generally, this helps when a player just wants to progress and discover the game without facing any friction.
 - **Desire for romping:** Similar items as for the emotion theme.

As you may notice, a large choice of items and areas of your game can be purchasable with money, and it is up to the data to guide you to find the best balance.

References

More information on gaming themes can be found in the following references:

- *The Hearts, Clubs, Diamonds, Spades: Players who suit MUDS series*, Richard Bartle, 1996
- *The Why we play games: Four keys to more emotion without story*, Nicole Lazzaro article, 2004
- *The 21st century game design book*, Chris Bateman, 2005
- *The Five ways games appeal to players* article, Jason Tocci, 2012, on the Gamasutra website
- *Motivations of Play in MMORPGs – Results from a Factor Analytic Approach* by Nick Yee

Summary

We have seen a wide range of topics in this chapter, but it was an important starting point because it can avoid some mistakes when analyzing the data. Imagine a situation where your game proposes a player versus player mode and a solo mode against artificial intelligence, and you wanted to evaluate which area of the game is more successful. You have collected information on the games made in each category and you might want to orient the development of your game according to the results. But as an analyst, you need to consider all of the possible meanings for the player; maybe your players play more in the solo area because they like it (desire for simple entertainment) but maybe they play it because the average gain of rewards is more appealing than in the player versus player mode (desire for reward). Thus, in order to avoid misinterpretations, you need to have a general understanding of game design and knowing situations described in this chapter will be definitely helpful.

In the next chapter, we will study the use of KPIs and evaluate which ones are the most important for the success of your game.

2

Common Key Performance Indicators

In this chapter we will see the main framework for running a game-analytics policy on a game: the notion of the **Key Performance Indicator (KPI)**. As we will explain it in detail later, these indicators constitute the basic structure for displaying the health of your game and its success in a simple way.

Definition and framework of Key Performance Indicators

KPIs are groups of signals, trends, which facilitate the decision process. They are supposed to be calculated on a regular basis in order to display the evolution of the game throughout time. They are usually presented through a dashboard, which means they are not supposed to be numerous.

Actually, they must not exceed a dozen in number because having too much information is a bad thing considering that KPIs are meant to synthesize the state of the game.

Criteria

The choice of those KPIs depends on several factors, and we are going to quote the main ones, which are as follows:

- The first criterion is the field of a KPI's application. This is the video game in our case, but this methodology is absolutely not exclusive to the video game field (web analytics, marketing, and business intelligence are also good fields of application).
- The second criterion is the simplicity of a KPI's meaning. KPIs need to be immediately understood by their readers.
- The third criterion is that the type of distribution chosen for the game is important, considering the video game field. KPIs will certainly not be identical between a Facebook browser-based game and a computer game sold on Steam.
- Finally, in specific cases, the type of game design intrinsic to the product may influence the choice of its KPIs, especially if the game has a particular game frequency (asynchronous) or an original progress curve (not based on levels, for example).

Another key point about KPIs is the temporality of their consultation, and actually there is no ready-to-use answer because it strongly depends on the temporality of your updates. If you are used to adding content and micro-features on a weekly basis, then you need to look at KPIs every week. On the other hand, if you add content on a monthly basis, it might be better to display the evolution of your KPIs every month. Notice that the notion of months, weeks, or days can be different from the calendar. If you add content on the 15th of every month, you need to harmonize the checking of KPIs with this date.

Now that those precautions have been highlighted, we are going to see what the most common frameworks for organizing typical video game KPIs are: these are the funnel acquisition, retention, and monetization.

Structure

Usually, KPIs of a typical free-to-play video game are organized through a three-step plan, which is called the funnel acquisition, retention, and monetization. Notice that for a game based on another model, for example, if it has a free limited version (demo) and a full paying version, this funnel needs to be adapted accordingly.

The category acquisition refers to the entry of new players into your game. It expresses the quality of your visibility on the market as well as the efficiency of your virality mechanics (for example, encouraging Facebook-sharing if you are a social game).

The retention is about the way people play your game, frequency, number of sessions, and also the loss of players at different levels of the game.

Finally, monetization is about the money you gain from your virtual store, in-app purchase (virtual goods), services, and advertising. Notice that KPIs linked to particular events outside of your game, such as a marketing program, should not be included in this sort of temporality and should be studied separately.

The following diagram highlights this structure, initially proposed by *Kontagent*:

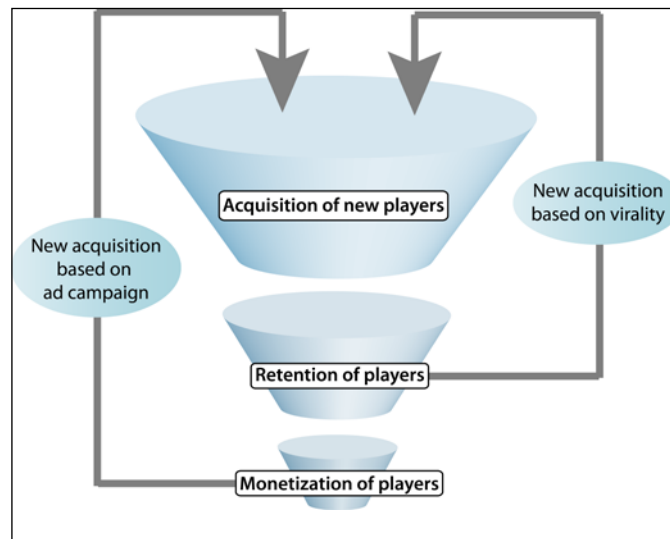


Illustration of the ARM funnel for video games

We are now going to see in detail KPIs for each category. For many of them, their temporality can change, and it is up to you to choose which one of the most relevant can be considered as your update pace. Do you need a temporality based on calendar or are you going to adopt a slippery temporality (where one month is equal to the last thirty days)?

Acquisition of new players

We will now see the most common KPIs in the first step of the funnel, the acquisition of new players:

- **New players:** This represents the gain of new unique players (player install), which allows you to calculate the growth of your game over time and the general quality of your acquisition. This can be defined by days, weeks, or months, slippery or not.

- **Cost per acquisition (CPA):** This represents the cost of acquisition for one player. This KPI is not so easy to evaluate because of the Facebook/mobile environment, where the source of traffic can be really diversified. The easiest way to do it is to adopt a one-month report pace, where you compare the total amount spent on marketing, Ad campaign, trailer, or press release during the month, and you divide this amount by the number of new players during the same month. Then, in order to use it as a decision tool, spend time to display all of the CPA by month on a chronological chart where you can see the evolution of the cost.
- **Source of traffic:** This represents the origin of new unique players (whether they come from Facebook sharing or an Ad campaign).
- **Tutorial loss:** This represents the loss of players between their first entry into the game and the end of the tutorial. Usually, the best way to proceed is to look at each step of the tutorial (depending on its duration) and identify which steps have a stronger loss than others. As a Key Performance Indicator, it is better to just take in considerations of its entry step and its end step. Another simpler variation of this KPI is the first day drop-off, which shows how many players have been leaving the game after their first day of play (what they have done during this first day is certainly the tutorial of the game).
- **Virality:** The virality is dedicated to the number of new players acquired, thanks to people already playing your game. Facebook-sharing is certainly the best example for this indicator, and it means that the player can send some invitations to play the game to his Facebook friends, and through this process, he will usually get some in-game rewards for himself.
- **K-Factor:** The K-Factor indicates, for one player, the average number of new players that will be brought to the game by him through Facebook-sharing or others networks, such as Twitter and Google Plus, or even in-game invitations generated by the game itself.

This overview will be completed by giving some magnitudes about these acquisition KPIs. It is particularly hard to find information about the range of KPIs according to each type of game. This is supposed to be private information, so you need to be cautious with any numbers that you may find on the Web. For now, we will give a general range of values that can be considered as a safe way to begin with.

According to the presentation of Mr *Michel Pierfritte* (Director of Game Analytics Research at Ubisoft) during the Web Game Conference 2013 hosted in Paris, we can classify the quality of a game acquisition especially with the first day drop-off variation:

- An outstanding or good game will have a first day drop-off variation not greater than 10 percent. It means that if only 10 percent of your incoming players leave your game after their first day, this is a really good sign.
- An average or OK game will have a first day drop-off variation between 10 percent and 33 percent of the incoming players. This range is good, but there is still some room for improvement.
- Finally, a risky game in terms of success will have a first day drop-off variation greater than 33 percent of the incoming players. If you experience a number of this sort, you need to dig into your data and look at what players are doing during their first entry in the game, and identify at which steps the loss becomes important. Notice that this metric can definitely be influenced by the type of your marketing acquisition campaign. If the promise that you have made with advertisements does not fit what the game contains, it is logical that a lot of players leave the game early.

You can find the full presentation of Mr *Michel Pierfritte* on Slideshare at <http://fr.slideshare.net/SylvainGauthier2/michel-pierfittedirector-of-game-analytics-research-at-ubisoft-modeling-of-players-activity>.

After acquisition, you need to evaluate KPIs for retention.

Retention of players

We will now see the most common KPIs in the second step of the funnel, the retention of players. This category is about how you keep your player involved into your game and how you retain them on a mid-term and long-term temporality.

- **Monthly active user:** This KPI indicates the number of unique players for one month.
- **Daily active user:** This KPI indicates the number of unique players per day.
- **Churn rate:** Also called **attrition rate** in other related fields, this KPI indicates the rotation rate of players. It is the ratio between players who have stopped playing the game and new arrivals. A high churn rate means that the fidelity of your players is weak, and if the quality of your acquisition decreases, your game might run out of steam very quickly. Usually, churn rate should be calculated on a monthly basis.

- **Average gaming time:** This KPI indicates the average number of hours spent by all of your players on a daily/weekly/monthly basis, depending on what you want to analyze.
- **Average number of game sessions:** This KPI indicates the average number of separated gaming sessions on a daily/weekly/monthly basis, depending on what you want to analyze. Close to the average gaming time, it is quite useful for game design when you want to adapt your gameplay to the temporality of the platform where your game is used. For example, a mobile game will certainly not have the same average number of sessions than a Facebook game.
- **Retention curve:** A large majority of games include a progression of the player's account, such as gaining levels. Exactly like the tutorial loss, it is usually very interesting to have a curve displaying the number of active players on each level of the game and see if there is a brutal loss on a specific spot. It helps to adjust the difficulty of the game and its progression curve.

Monetization of players

We will now see the most common KPIs in the third step of the funnel, the monetization of players. It is important to notice that monetization remains the overriding aspect of a free-to-play game because it represents the main source of income. Having a good acquisition as well as a good retention becomes useless if the monetization does not follow, and it is important to keep this in mind while deciding which aspect of the game efforts must be focused on.

- **Paying users:** This KPI shows the percentage of players who have made at least one purchase inside the game compared to the total number of players. You can define this KPI with the date of the purchase, and choose to show only players who have made a purchase within the last day, the last week, or the last month.
- **Average revenue per paying users:** This KPI shows the average amount of real money spent into the game for one paying user. Similar to the paying users' KPI, you can define it for one day, one week, or one month.
- **Average revenue per users:** This KPI shows the amount of money gained for one user (paying or not).
- **Lifetime value:** This KPI is supposed to show the total amount of revenue generated by one user considering the current average revenue per the paying users' KPI of your game and the retention expectancy. As we will see later, this KPI is hard to calculate because it is always evolving, but it remains very important in order to estimate if an acquisition campaign is worth its cost or not.

- **Conversion rate:** This KPI shows the number of non-paying users who have become paying users. It is very useful for measuring the impact of a feature, especially if you can define this conversion rate for each level of the game.
- **Source of income:** From a free-to-play point of view, the source of income can be very diversified, and it is noticeably interesting to evaluate which area generates the best income for the game. This indicator will have categories, such as income from virtual items, Ad revenue, and branding revenue eventually.
- **Items gain:** This KPI shows the gain from every purchasable item of the game. In order to use this metric, you need to classify those items: additional content, acceleration of the progression, visual appearance, and upgrade of power are the most common categories for free-to-play items. The previous chapter has given you a good overview of how you can structure purchasable items through gaming themes.

Working regularly with KPIs

As we already said, KPIs and their framework are not supposed to explain phenomenons. They give a simple and quick overview of your game. From this statement, they are also very useful if you want to calculate the long term profitability of your game.

A common operation is to compare the lifetime value KPI with the cost per acquisition. If the lifetime value KPI is superior to the cost per acquisition KPI by a large amount, you might want to focus your strategy on the acquisition of new players. On the other hand, if the cost per acquisition KPI is too high, or if the lifetime value KPI is too low, you will need to increase (or decrease) their value.

Generally speaking, KPIs provide a general overview that can help you to decide where you need to focus your efforts on:

- If short-term acquisition has a problem, you might want to look at your tutorial stage to check if the gameplay is not explained properly, is it too long, or at which stage of the tutorial players are leaving the game
- If mid-term retention has a problem, you might want to look at the progress curve of your game to check if the game is too hard at a certain stage, or is it too easy so that it becomes boring, or are resources gained by your players significant enough so that they can have rewards regularly and stay into the flow

- If monetization has a problem, you might want to look at the balancing of your virtual economy and the type of items bought, what kind of advantage the player received from buying and is it relevant for them, or are prices too high
- If your game has two different sorts of money (one gained through in-game action and another gained with real money), what is the balancing between both

In the end, making a habit to look at those KPIs will definitely increase your reactivity and the efficiency of the changes you will make on your game. They are a great tool and do not need advanced statistical skills to be used. From this point of view, even small studios can greatly benefit from them. Moreover, do not limit yourself to a particular number of KPIs. Their number needs to change from one game to another and according to your objectives. There is not a fixed set of KPIs for everything.

Summary

This chapter has shown that using Key Performance Indicators can greatly help you to understand what's happening in your game, from the acquisition of new players and their retention inside your game to their monetization. Moreover, they will especially be useful to evaluate which of those areas you need to focus on.

Now that we have finished seeing the general framework of Key Performance Indicators and how they are supposed to be used, we will evaluate the different tools proposed by the technology and the market, which are dedicated for statistical video game analysis. Before analyzing KPIs, you need to collect data and display it, and the next chapter will highlight this underestimated aspect of the field.

3

Environment and Tools for Data Analysis

In this chapter, we are going to see the most popular tools for data analysis, from data collection to its exploitation. We will explore different solutions, including storage solutions and free tools to paying software, and we will try to give a good overview of what they proposed and what are their advantages.

Typical programming environment for data mining and storage

In this first part we will have a general look at the most common development environments for data analysis, which are MySQL and NoSQL (with Hadoop essentially for data analysis).

MySQL

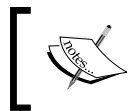
SQL (Structured Query Language) is currently one of the world's most widely used languages for database management. It implies a relational structure with data being structured into tables with columns, and it has been developed in order to promote connections between those multiple tables; from this point of view, SQL is more efficient when applied on databases that are already fixed and structured than on large, fluctuating databases with regular updates on their foundations.

Through queries and joins, you are able to retrieve, filter, and calculate data with a lot of freedom. Notice that while SQL refers to the language used, storage solutions based on SQL are numerous: MySQL, Oracle, Microsoft SQL server, Microsoft Access, and the list goes on. Currently, the only open source solution is MySQL, and you can download it for free on the Web.

Another important feature in SQL is the server-replication facility. One server can be duplicated and have several identical slaves. This means that if one server goes down, the data can be accessed through another clone. That being said, the process of managing different servers still requires some work on the part of the developer. When a slave goes down, all of the requests are sent to its master, and it might sometimes generate too much flood for it if the data is important. On the other hand, if the master server goes down, it is necessary to configure one of its slaves to become a parent and be updated in its structure.

Very simply, managing a database structure that you have built on your own means you will need at least one person with the appropriate skills to work on it when required.

SQL is essentially used on websites (coupled with web frameworks such as PHP, thanks to its very low price and its facility of installation). Generally, it is the best tool to use for fixed databases with no important scaling, and from this point of view, it might not always be the best for video game data mining, where flood can become very significant within a few days if a game is a important success.



Thanks to the maturity of SQL and its libraries, scalability is definitely not so bad, and in the end a lot of things depend on your objectives and what you want to do with your data.

We will now have a look at other new environments belonging to the NoSQL category.

NoSQL

NoSQL (NotOnlySQL) differs fundamentally from MySQL because it is not based on the relational model. NoSQL databases have grown up in the last few years, thanks to Google (Big Table) and Amazon.

One of the first arguments in support of NoSQL is the better scalability it has. In MySQL, typical joins in multiple relational databases can really slow down the system, and when experiencing a very high peak attendance, your servers might simply crash.

Without going into further details, as this is not the main subject of the book, we can point out that NoSQL is better at enduring peaks and has better scalability in general. Initially invented for huge databases that were fed by web traffic, this solution has also emerged as a very good choice for online video games because it is common that they can generate great traffic as well.

Nowadays, the debate between SQL and NoSQL is quite senseless because they address different uses and functions with data. A hybrid solution has also emerged and appears very efficient, where the NoSQL structure is used for the storage aspect while the request aspect remains based on SQL. Relative to this, the Hive framework is an interesting evolution of Hadoop because you can have a non-relational database (Hadoop) that you can access and manage with the language HiveQL, which is very similar to SQL.

Hadoop and Hive

Strongly inspired by the **MapReduce** system of Google, Hadoop is a framework in Java that has been built to deal with a very large amount of data, such as petabytes. This tool becomes very interesting when you want to deal with millions of monthly active players.

Its structure includes a MapReduce engine and the **Hadoop Distributed File System (HDFS)**. Contrary to a relational system, where all of the data is structured into tables with columns, the data is stored in a disorderly manner and you will structure it through the use of key-value pair. If you add the Hive framework onto Hadoop, you will be able to use a SQL-friendly language that will be translated appropriately for MapReduce tasks. Note that even though HiveQL is similar to SQL, what happens in queries is radically different from what happens on relational databases.

What you need to remember is that this solution will be the best if you have games that generate very large amount of data. This large amount can come from the detail of events that you are collecting from your game (if you want to collect everything, for example) or simply if you have lots of players in your game (one million monthly active users, for example).

Moreover, this tool will be a very good choice if you are planning to apply complex statistical treatment; in other words, if you want to perform advanced analysis on a very large amount of data.

Tools available on the market for quick data mining

Now that we have presented the storage environment, we will explore readymade tools and their interfaces that allow people that are not developers to explore the data without using SQL or any other tool.

Available free tools

In this section we will see two main tools for basic data analysis, namely Facebook Insights and Google Analytics. The interesting thing about them is their free access. Although they do not provide a user-centric tool, they remain interesting as a first step for beginners.

Facebook Insights

Facebook Insights provides you with metrics about your content; it can be a fan page as well as an application (like a game). Here you will find general trends about usage and consumption and demographic information (country, city, sex, and so on) about your sample. This information will be available through **Insights Dashboard**. The following is an overview of this dashboard:



This dashboard can be used for a fan page or application, but it can also be used with any website if it is associated with a Facebook app, or even with your own user account.

Secondly, Facebook also proposed **Open Graph Insights**, which allows you to go further into the analysis of your content, and is dedicated to applications only. Here are the main metrics covered by Open Graph Insights:

- **Site engagement:** This section displays an overview of your application and is essentially focused on the acquisition of traffic. It starts from your published actions, to the number of people reached (impressions), to the number of referrals (number of clicks).

- **Story impressions and clicks:** This section is a break down displayed on a timeline, which shows the click rate (impressions versus referrals) for any given action.
- **Story likes and comments:** Similar to the previous section, this one displays the like rate and the comment rate for any given action on a timeline.
- **Timeline unit activity:** Users can highlight, hide, star, unstar, or even add an application to their profile. This section shows these actions on a timeline and you can break down the results by actions.
- **Demographics:** This section classifies your players by gender, source, language, and country. You can filter these classifications according to particular actions, such as story likes or timeline unit impressions.

As we can see, options look numerous but one of the problems related to Facebook Insights is certainly about the choice of having a **page-centric** tool that prevents you from doing advanced analysis on your game (especially if you want to have a classification of users according to specific actions inside your game). For example, you cannot have a list of user IDs associated with events inside the game; you only have access to global numbers on specific events proposed by the tool. This is the main problem of Facebook Insights, but you must understand that this is a free version.

Google Analytics

Google Analytics is a free service as long as your audience count does not go beyond five million page views by month coupled with AdWords (which is an advertising service); after this amount, the service becomes payable.

Essentially dedicated to e-marketing and websites, Google Analytics can also be used for browser-based games, Facebook, or mobile applications.

Like Facebook Insights, Google Analytics proposes a page-centric framework for collecting and analyzing data. Thanks to the dashboard, you will be able to segment your traffic according to particular events (like paying events) or create your own events with a specific tool. You can display those results on a timeline and compare them with past results quite easily. A few statistical tools (such as pivot tables and filtering tools) and export functionalities (CSV, Excel, or PDF) are available too. Moreover, you can configure an alert if your traffic is experiencing some strange evolutions. Finally, you also have access to data-visualization tools, such as animated charts and funnels.

All in all, this tool remains very useful when compared to its price (free in many cases), and exactly like Facebook Insights, the problem is that there are no user-centric collection possibilities.

We will now have a look at solutions that can realize this user-centric data collection.

Commercial solutions

In this chapter, we will see three commercial solutions for data analysis. It is important to note that far more solutions exist on the market. However, the chapter would not end if we had to describe them all. Therefore, examples here are typical examples of what you will usually find on the market when considering prices and quality of services. Kontagent can be considered as an expansive but high-quality service tool, whereas Honey Track belongs to the less expansive and scalable tool category. However, it might require more development efforts from your side (especially the integration of their solution). Finally, the third example will be about a mobile-focused solution like Flurry, which proposes an easy ad-integration on your application.

Kontagent

Kontagent was founded in 2007 and has worked with world-famous game developers such as Electronic Arts and Ubisoft. As mentioned previously, this option belongs to the category of solutions that are quite expensive but ensure a high quality of service.

Their analytical solution can be divided into three parts, namely metrics, data science, and consultation. They use a time-tested platform that delivers real-time analytics as well as insights, such as session lengths and player performance, everything KPI related, as we have seen in the previous chapter. The advantage of this solution, outside of having consulting and good support, is definitely the fact that you can have reports that are user centric (you have your own options of filtering, plus you know the user ID, which allows you to make junctions between behaviors and users actions).

Recently, Kontagent has been more and more focused on the mobile field, and similar products are solutions like game analytics.

Honey Tracks

While Kontagent has a large catalogue that is not exclusive to video games, Honey Tracks is clearly a video game-dedicated solution. Its price depends on the amount of flux your game is generating, and you need to integrate their solution by yourself inside your game. Once this is done, you will have access to your data and events through a smart graphic interface that will facilitate the collection of the data [for example, you will not need to process MySQL requests (or whatever language you use) in order to collect the data], and from this point of view, there is a real gain of time.

Through Honey Tracks' framework, you will be able to perform user-centric analysis based on your own events and others things such as cohort analysis. On top of that, you will have many data-visualization options for displaying your KPIs.

The scalability of its price makes this solution clearly interesting for small and medium studios that want to start deeper data analysis.

Similar solutions are products such as Kiss Metrics, Lvlboost, or Playnomics.

Flurry Analytics

Flurry Analytics is clearly dedicated to mobile analytics. It is quite easy to integrate and offers a dashboard for common KPIs. Actually, the real advantage of this solution is about the ad-pushing provided with the analytics framework. You have many options to orient your customer acquisition and you can measure the effectiveness of your campaign quite easily. From this point of view, calculating **Return Of Investment (ROI)** is easy.

On the other hand, the analytics part of the solution is less developed; the approach here is clearly marketing oriented.

Tools available for analysis

In this section, we will see the different tools available for pure statistical analysis. Even though some of these tools include methods of collecting data inside databases, exactly like solutions we described in the previous section, the idea here is to be focused on the quality of analysis methods proposed and their visualization options coupled with the price of the software. We will also give a few observations about their difficulty of mastery. Note that once again, all of the tools available are not covered; the idea is to propose a quick overview.

Open source tools

In this section we will see some of the popular open source tools for analyzing and modeling data.

R-Project

R-Project is primarily a mathematic environment dedicated to statistical treatment and data analysis. It is an open source tool based on the S language. R environment is very interesting for several reasons:

- It is free
- Its possibilities in terms of statistical methods are large in number when you use its extensions (far more than a hundred)
- Once you master it, you can develop your own tools and your own process with plenty of freedom

This last statement leads us to its major drawback: the software is hard to master and requires writing instructions using R language in order to realize analysis. Contrary to other tools that we will see later, you do not have any user friendly interfaces that can facilitate the choice of methods and algorithms.

Gephi

Gephi is an open source software dedicated to network analysis and visualization. It is well known for displaying the network and connections between elements through smart graphs and charts. Close to a Geographic Information System, this software will be very interesting if you are looking for advanced data-visualization options.

Commercial tools

Commercial tools are supposed to propose more advanced options than free open source tools, especially regarding ease of use.

SPSS

Statistical Package for the Social Sciences (SPSS) was one of the first programs dedicated to data analysis. Acquired by IBM in 2009, the tool has evolved and is now targeting a larger audience, from finance to video game analysis.

It offers a wide range of several statistical tools that we will develop in *Chapter 6, Data Visualization*, and those methods can be accessed through its GUI (drop-down menu) as well as instructions using its own language (named 4GL). From this point of view, the software combines the freedom of R (where you can also write your instructions freely) with the handy aspect of having a GUI.

Statistica

Exactly like SPSS, Statistica proposes a framework for data analysis and has its own GUI. It is edited by the StatSoft society and proposes more than sixty modules (equivalent to R extensions) that cover a wide range of statistical tools, from linear regression to variance analysis and segmentation for predictive analysis.

As for the other tools, you will find many options for visualizing data and the interface ensures you write minimal instructions.

Rapid Miner

Rapid Miner is similar to previous examples, except that it can also be provided with a data collection service linked to your database. In other words, you will be able to perform both data collection and data analysis with one tool instead of two. This particularity leads us to another one, which is the way the product is marketed. One version with all of the analysis tools is available for free, but you must pay if you want to use the software on your own database directly and get technical support and consultation.

Rapid Miner is interesting because you can definitely take your time to test and experience it a long time before choosing to buy the entire solution. From this point of view, it combines the advantage of an open source tool as well as a commercial tool.

The following screenshot shows its GUI and an example of advanced data visualization:

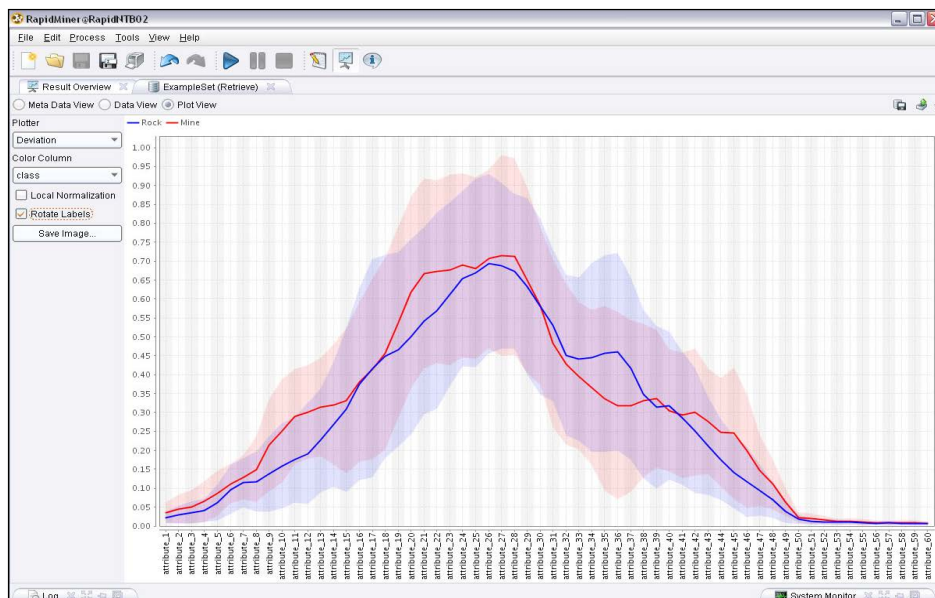


Tableau Desktop

Tableau Desktop is a software that is exclusively dedicated to data visualization and has the advantage of ease of use and provides a large number of options for data visualization.

If you are planning to limit yourself to the study of KPIs, this software is definitely the best you can invest on.

With many possibilities to connect the software to your database (you can connect it to a Hadoop/Hive database, for example), you will be able to manipulate and display your data on charts and diagrams through a drag-and-drop system that is really user friendly.

On the other hand, possibilities of complex statistical operations are less important than on the other tools previously presented.

Summary

In this chapter we have seen a wide range of tools along with their environments (SQL or NoSQL). The important takeaway is that integrated tools (linked to a particular environment, such as Facebook Insights or Google Analytics) will generally offer you ready-to-use analytics that are not user centric. On the contrary, paying tools will take you to the next level with user-centric analysis, allowing you to apply advanced statistical treatment on them. Free open source tools for analysis usually have the caveat of being hard to master (R environment is a good example), but keep in mind that you can do the same thing as you can on commercial tools once you master them.

In the next chapter we will see a particular aspect of game analytics – the generation of content based on metrics.

4

Game Analytics and Generation of Content

We have seen the typical KPI used for online video games and different tools available for collecting data in the previous chapters. Before exploring the analysis aspect and the statistical tools in depth, we will detail a particular aspect of game analytics, that is, the generation of content and the A/B testing.

The *Testing* section will highlight two types of testing, the A/B testing (which includes only two different features) and the multivariate testing (which includes several features to test).

Testing

Historically, the A/B and multivariate testing methods existed in other industrial fields before being applied to online video games. They have made their first appearance in the digital field with the rise of web analytics. Logically, those methods have also started to be used for online video games.

The A/B testing

The concept of the A/B testing is quite simple:

- You have two versions of one feature (A and B), and one metric which measures the success of the test. Usually, it is going to be a click on a particular button, such as buying coins, accepting to install the game.

- Usually, those features will be subjected at the same time to a sample of your audience (notice that the sample must be statistically significant). The sample is going to be chosen in order to fit the objective of the test. If the test is about improving Facebook virality for new users, you will not include in your sample old users for whom virality potential is already tried out.
- Once the criteria selection for the sample is done, features tested are distributed randomly between your targeted audience.

At the end, you pick the version with the best success and apply it to the entire game.

The multivariate testing

Now that we are aware of the basics of the A/B testing, it remains exactly the same for the multivariate testing, except that the number of features tested goes above two. You can actually have a high amount of features tested at the same time. The only constraint being that you will multiply the number of versions and so the size of your sample must increase accordingly in order to keep your test statistically significant.

If the features can be combined between them, results for the multivariate testing will be displayed in a matrix with all of the versions tested. For example, with three different colors for a button, three different pictures, and three different amounts of purchasable coins for one pop up, the number of different pop-up versions for this test could vary between three and twenty-seven (where all combinations are tested).

Recommendations for good practices

Since live testing has a cost, it is important to choose carefully what is going to be tested. The most common areas for testing in the social/mobile sphere are about the conversion rate for the acquisition (for example, first page inviting the player to start the game and eventually the tutorial phase), and also the monetization (the virtual shop interface, and also incentive pop up for discount).

Depending on your game, keep in mind that other areas can be a very good place for testing (important game design turning points, virality...), the preceding examples are just typical areas.

Following those precautions, what exactly is being tested with this method? We will develop this part with an example later in the chapter, but the following are a few guidelines of what can be tested:

- The call to action element (usually a button)
You can test its wording, color, shape, and even placement.

- The text content
If you have some descriptive text, which explains a few things on the action or on your game, you might want to test different sentences and descriptions. You can also test the entire text on your page.
- The general graphic style
Layout, choices of design, choices of graphic representation, choices of picture shown to the player.
- The choice of product pricing and its integration into your economy
Paying players can have different behaviors; they can be impulsive but they can also have a general strategy with their credit consumption so you want to know what are the best prices for your audience.

Once you know what you are going to test, you need to keep in mind that the relevancy of your test will improve if you are strict with the process. The synchronous aspect, for example, must be respected. Otherwise, time-dependent facts (such as a holiday period happening only for the first sample tested) may influence the results of the study.

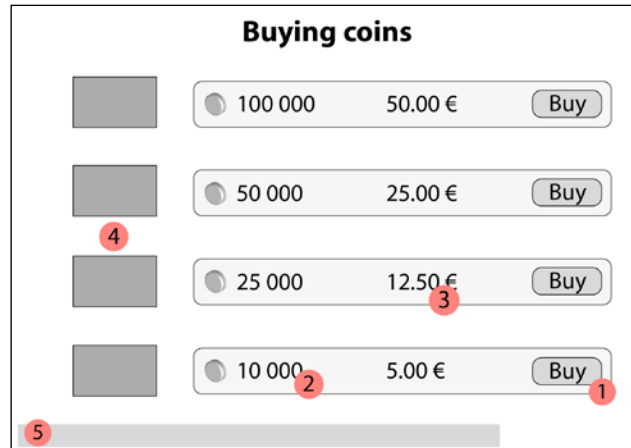
Following this statement, you must also evaluate how much time you need to run the test. Common time durations for a test are usually between one week and two months. Actually, the amount of time you need to run a test totally depends on the size of your audience. If you have a game with millions of players, you will reach your objective far more quickly than if you have a few thousands of players.

At the end, it is important to notice that many A/B tests sometimes just do not have any relevant results. The key is to practice a lot and to be persistent, maybe half of your tests will fail, but if one increases your income by a large amount, it is worth the case. Do not forget that this tool comes from the web field, and it might need some adjustments considering you are trying to apply it on video games. Make your own recipe, make a focus group on what should be tested and include your team (especially game designers) in the process.

We are now going to look at an example case, and the idea will be to illustrate how many things can be tested on a single monetization pop up.

Case study – monetization pop up

There are two steps in this part. First, we will try to describe all of the features that can be tested considering the following screenshot. Secondly, we will see what different versions of this pop up could be, and what we can expect from the test. The objective, of course, is to have the highest number of clicks on the **Buy** button.



As we can see, the preceding screenshot shows a typical incentive pop up proposing to buy coins usable inside the game (for buying items or unlocking content, for example).

Numbers inside the red circle represent the area of the features which are going to be tested. Normally, the choice of these features is subjected to a brainstorming session within your team, because you need to define and anticipate what the most relevant features to test are. In this case, we will simply list all of the possibilities available for testing, and then we will propose two versions of the pop up.

The next section will highlight all of the possibilities available for testing, and will be classified according to each red rounded number located in the preceding screenshot.

Inventory of each feature

Linked to the red circle marked 1 in the preceding screenshot, we can test the following things on this feature:

- Color (red, green...)
- Style (writing)
- Size (bigger or smaller)

- Special effects, such as a highlighting effect on the button
- Amount of coins proposed

These aspects are often underestimated but they remain very important. There are lots of strategies possible here and you need to take into account the game design of your product, especially the **friction** side (the "friction" term here refers to the frequency with which a player in the game becomes blocked). Even if you do not have a lot of friction in your game, you need to consider the most beneficial one for you, many small purchases from a large audience or a few high purchases from a very few people (the so called "whales"). The following are a few precepts:

- As we can see, the pop up presents four options. Whatever the exact number is, the most important options are the first one (the cheapest) and the last one (the most expensive).
- The first option might be used as a way to toggle a non-paying user status towards a paying user status. Many marketing researches have proven it. Once the first act of paying is done (even if it is one cent), the chances of having a second one increases significantly. Usually, you will use an amount that fits the current short-term objectives of your player, but be aware that those short-term objectives certainly vary from a level to another. Quests and objectives are not the same at level one as they are at level ten (or you have a problem in your game). Therefore, adapt the offer to its temporality, do not force a player to buy the category higher because he lacks one coin with the cheapest offer for his objective.
- The last option usually has the most interesting value for money and you need to balance it with your in-game economy, and find an adequate amount for your very engaged players. It needs to be adequate because it fits their long-term objectives and it's adequate for you because it is not too generous (or you will lose money). The balance is hard to find and that's why simultaneous testing is so much interesting.
- Price of coins in Euros

This feature is about the Euro conversion rate of your offer. Usually, you will have a better value for money as long as the offer is higher but you can test several things. For example, maybe the cheapest offer made for the conversion should also have a good value for money, you can choose an amount, which is not rounded in order to make the offer look less expensive. At the end, you can also inform the user that the last offer has the best value for money thanks to a symbol aside.

- Picture(s) displayed beside the offer
Again, this feature can be very interesting to test and it is sometimes underestimated. The following are a few proposals:
 - You can show a valuable item that looks more appealing as the offer becomes bigger. You may want to awake the desire of something valuable and precious.
 - You can show one of the most famous characters of your game in order to capitalize on the affective link generated by the character for your players.
 - You can show some items that your player will be able to buy with those coins. It might be interesting if you want your player to have a clear view of what the amount means exactly to the character.
- There are many other options to test, which are not especially indicating the current picture. Some are as follows:
 - Adding a reassuring padlock symbol with the name of the service used at the bottom of the page
 - Adding optional incentive information, such as **the money gained will improve the quality of the game**
 - The overall layout of the page
 - The overall graphic design of the page

Concrete examples of versioning

Now that we have looked at a large variety of options, it will be interesting to see different examples of designs for that kind of pop up. This is a bit of bonus part where the idea is to make the reader feel the difference in terms of perception. As you will notice, I have voluntarily chosen to create opposite examples.

First example

Have a glimpse at the following screenshot and try to put yourself into the situation of a player:



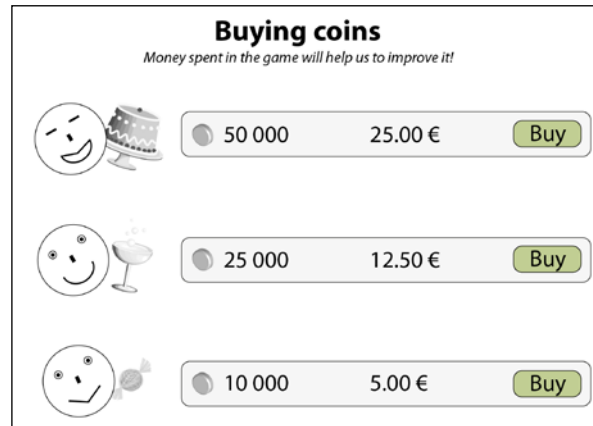
Many statements will follow the remaining hypothesis and will not really have a strong value outside the pedagogical aspect of the exercise.

So as we can see, this style of pop up could fit a systemic game like those inspired by slot machine (casino). The choice of displaying five different choices is to make the highest one more interesting due to the better value for money at each rank. Pictures displayed on the left-hand side of the page are supposed to be a treasure, which becomes more appealing for each offer, promoting again the highest choice for the user. Considering previous elements, I have chosen to set the color of the button to yellow (the color of gold), and a mini symbol informs the user about the best value for money. Finally, since I tend to promote a substantial purchase instead of a small one, I want to improve the feeling of security and so a padlock symbol is displayed at the lower-left.

Generally speaking, the design of this pop up would fit an audience where monetization relies on a minority of paying users spending high quantities of money on the game (the so-called "whales"), but the only way to be sure of this statement would be to test it.

Second example

Now, have a glimpse at the second screenshot:



This example is clearly more "traditional video game" friendly. Only three choices are available for the player, and pictures aside are supposed to represent a cheerful character of the game associated with a purchasable object. The idea here is to promote a user-friendly design where the affective link of the player with the game is emphasized and it also indicates approximately what sort of objects might be bought with those amounts. Finally, a sentence below the title of the page reminds us that money gained in the process will serve the development of the game for the future.

Globally, this sort of design could fit a game focused on a particular audience (a "niche") quite involved into the game and where monetization relies on multiple small purchases. But again, only testing would verify that sort of statement.

Summary

This chapter has provided an in-depth presentation of how metrics can help to generate content. What is important to notice is that those methods are useful once you have an idea of what is interesting to test, and this was exactly the purpose of the second part of the chapter. Notice that the A/B testing, exactly like KPIs, do not give you the reason why the hypothesis A is better than the hypothesis B, it just gives you the best hypothesis according to an objective.

In the next chapter, we will see advanced statistical methods that can be applied on the video game field.

5

Advanced Analysis and Statistical Methods

In this chapter, we will have a look at the tools for advanced analysis, essentially for statistical methods. As many people might be reluctant to learn about this particular field, the idea will be to present most of the methods without using any mathematical formulas, and most of them will be illustrated with an example from the video game field and a concrete game-design situation.

Note that all of the tools listed in this chapter do not represent the entire domain. Instead, this is a general overview which is not comprehensive.

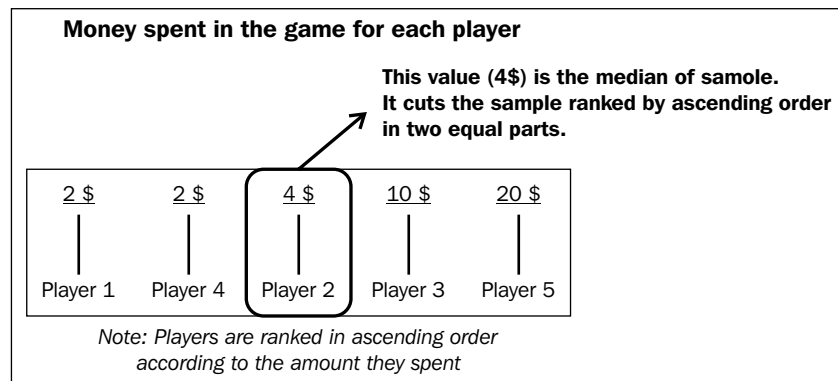
General statistical description

In this section, we will highlight the basic notions of statistics, such as the average and the standard deviation. We will also explain some of the most important statistical laws (like the normal distribution) in brief.

Central tendencies

The following indicators tend to highlight the central point of a sample. **Central point** refers to the visualization of a statistical sample on a chart.

- **The arithmetic mean (or average):** The mean is certainly one of the most used indicators in any statistical study. It represents the general tendency of one population. For example, if you have five players who have spent two, two, four, ten, and twelve dollars respectively on your video game, the average is equal to the total money spent by each player divided by the number of players in your sample. According to this example, an average of six dollars was spent on the game. Note that the average clearly has one major default—it can hide disparate tendencies. For example, a set of five players who have spent zero, zero, zero, zero, and thirty dollars respectively on your video game has the exact same average result of the previous example: the mean of this sample is also six dollars. As you can notice, the mean tends to have a certain degree of relevance in the first example. However, in the second example, the value of six totally hides the reality of the sample, where finally only one user spends money on the game.
- **Median:** This value is close to the average as it cuts a sample in two equal parts. Considering the previous example, if you have a sample of five players who have spent two, two, four, ten, and twelve dollars respectively on your video game, the median of this sample is four. The following diagram illustrates the example of this median:



As you can see, indicators that are frequently used for reporting **Key Performance Indicators** must be considered with caution because they show a general tendency that may lack precision in order to explain the dispersion of a sample.

Dispersion tendencies

Opposite and complementary to central tendency indicators, indicators for dispersion try to highlight the level of difference inside a sample. Keep in mind that if the following indicators are low, it means that the sample is quite similar, and the indicators may finally also be used to show homogeneity inside a sample:

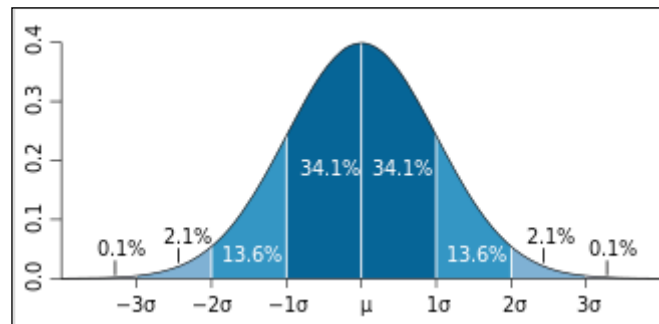
- **Standard deviation:** The standard deviation represents the average of the distance between every object of the sample and the arithmetic mean of the entire sample. Considering the first example again, where five players spend two, two, four, ten, and twelve dollars respectively on a game, the average distance between them and the mean (six dollars) is four. If we apply this to the second example with only one player spending thirty dollars on the game, the standard deviation reaches nine and a half. It is usually defined and calculated as the square root of the variance.
- **Variance:** Close to the standard deviation, the variance also shows how far a set of numbers is spread out. The main difference is that the number can be negative because it is the squared deviation from the mean.
- **Interquartile:** As opposed to the standard deviation and variance, interquartile is composed of two values that cut one sample into four equal parts that are located at each extremity. Consider the following numbers: 2, 3, 5, 10, 12, 13, and 25. Their median is 10, and so their interquartile values are 3 and 13.

Dispersion indicators by themselves are quite interesting, but as we will see later, they are very useful for building more complex analysis; for example, segmentation and clustering.

Statistical distribution and laws

A law in the statistical field is supposed to describe a given behavior for a set of numbers. Initially, statistical laws and distributions appeared with the study of gambling games. The motivation behind those studies was to improve the chances of winning. Concretely speaking, laws are a theoretical assumption, an expectation of how structured a set of numbers should be.

- **The normal distribution (also called Gaussian distribution):** It corresponds to a continuous probability distribution that takes the form of a **bell curve** when all of the data points are displayed on a chart. Concretely speaking, the center will contain the highest number of a value and correspond to the highest point of the bell curve, whereas values on extremities decrease quickly on either side. The probability factors of a normal distribution are identified through the following rule: 68 percent of the distribution must be included in the area that is covered by one standard deviation, 95 percent by two standard deviations, and 99.7 percent by three standard deviations. What is important to understand is the fact that if a distribution is normal, lots of analytical processes and operations can be applied to it (by a data analyst). This is not possible if the distribution is not normal. The following diagram will help you understand this distribution:



Source: http://en.wikipedia.org/wiki/Normal_distribution

As you can see, in a normal distribution, 68 percent of the sample is supposed to be included in the first two standard deviations.

- **The student distribution:** Following the normal distribution, the student distribution and its test belong to the family of continuous probability distributions and are especially used for estimating the mean of a sample when the standard deviation is unknown. This process is essentially useful for other advanced analysis treatments, such as for the linear regression analysis.

- **The Poisson distribution:** This distribution corresponds to a discrete probability distribution and is focused on the rhythm in which an occurrence happened with a regular frequency. Initially, this law was introduced by *Simeon Denis Poisson*, who studied the frequency of judgments in the civil and criminal field. Concretely speaking, it is supposed to predict the degree of spread around a known average rate of occurrence (source: *Statistics – The Poisson Distribution, Umass.edu 2007*).

Correlation and regression between variables

We will now review an important aspect of advanced analysis methods. It is about the correlation between variables. Correlation is a very important topic because it allows you to study how a given factor (let's say, for example, gaming frequency time) impacts another one (for example, how much money do players spend on the game).

Before starting, we need to define the term's correlation and regression as they are used in the statistical field:

- **Correlation:** It refers to a symmetrical link between two variables, but there is no dependency between those variables – one cannot predict the other one. For example, consider the sales of ice cream and sunglasses during a year. Both are linked because when one of them starts to increase, the other one increases as well, but one cannot explain the other one because the explanatory factor is the weather.
- **Regression:** On the other hand, regression refers to an asymmetrical link between two variables, where one variable can explain the other one but the opposite is false. Considering the previous example, the weather conditions and the sale of sunglasses belong to this category.

We will now have a look to the types of variables used in the statistical field.

Types of variables

Before studying some of the methods focused on correlation and regression, it is important to define the types of variables that are used simply because most of the methods work only with a certain type of variable. Usually, each sort of variable has its own set of methods, even if some of the methods are able to compare and analyze the correlation of two different kinds of variables.

- **Discrete variable (or integer number):** This refers to an integer number. For example, the number of children in a family cannot be two and a half – it needs to be an integer.

- **Continuous variable (or non-integer number):** This refers to a number that can have a finite or infinite number of decimal. For example, the weight or size of a person.
- **Binomial variable (yes or no):** This refers to a variable where the value is either 1 or 0, which usually means false or true. For example, a binomial variable in the video game field could be the paying status of a user: 1 would mean that the user is a paying user, 0 would mean that he is a non-paying user.
- **Qualitative (or categorical, or multinomial) variable:** This refers to a variable where the value is a finite number of categories, which will usually be translated into numbers. For example, you could have a set of players with the following variables: 1 if the player is an adult male, 2 if the player is an adult female, and 3 if the player is a child. This sort of variable can also be used for translating answers from a survey, where each number corresponds to the different choices available for each question.
- **Ordinal qualitative variable:** This is similar to a qualitative variable. The only difference is that categories have a quantitative signification. For example, when the satisfaction of a user is noted on a scale where 1 means low satisfaction and 5 means high satisfaction.

There are other categories and sub-categories of variables, but those are the principles. We will now explore some of the most common methods for identifying the correlation between variables.

Chi-squared test

The chi-squared test has many different uses, such as testing the suitability of a given data set for a particular distribution (explained in the previous section), but here we will just consider its use for the analysis of independency between two variables.

A typical way to proceed is to formulate a null hypothesis that is going to be verified through the test itself. The idea is to measure the gap between the real distribution and a theoretical distribution where the dependency is absolutely null. The gap observed will allow us to determine if the null hypothesis can be accepted or not.

One of the common uses of the chi-squared test is to apply it to the categorical variables that will be displayed in a contingency table. In simple words, a contingency table shows the results from each sample according to each variable as well as the total observed for each of the rows and lines of the table.

The theoretical table is calculated, thanks to the first contingency table, and a comparison is made in order to evaluate the distance between both distributions.

A few important points about the chi-squared test:

- This method does not allow you to identify the symmetry of the dependency between two variables. (Does variable X influence Y or is it the opposite?)
- There is no ready-made value that allows you to easily conclude if the distance between the theoretical distribution and the real one is relevant. This value varies from one case to another, and it is up to the data analyst to determine if the test is relevant or not.
- The test always has an error margin equal to 5 percent (this value is arbitrary and is the most used value in general).
- Generally, we can point out that the chi-squared test will be interesting as the first explorative step for quickly identifying if certain variables constitute an interesting topic to dig in.

We will now evaluate another method called linear regression.

Linear regression

Linear regression is supposed to model the regression between two or more variables that will be called Y (the dependent variable) and X (the explanatory variable(s)) respectively. This method is very interesting in order to quantify the strength of the link between the dependent variable and its explanatory variable(s).

As compared to the chi-squared test, this method allows us to measure not only the strength of the link between variables, but also the symmetry of this dependency at the same time, making this method more oriented towards a conclusion rather than exploration.

Developing such models is the first step towards the prediction. When you know the level of relationship between X and Y, if a new X is added, you can predict its Y value, thanks to your previous model. For example, if you have realized a study on your player database with `money spent in the game` as the dependent variable, and have several explanatory variables such as `gender`, `age`, or `source of acquisition`, you will be able to predict approximately when a player enters your game and how much money he is going to spend according to your model, especially if chosen variables are shown to be relevant by the linear regression.

Of course, such a model is never perfect and must be reworked and refined all of the time. If your game is updated, or if you change some of the features in it, you need to look again at the data in order to update the model accordingly, and eventually change some of the explanatory variables that are used.

Concretely, the idea is to first verify that the dependent variable has a linear relationship with the explanatory variable. By linear, it very simply means that the scatter of data on a chart may be synthesized by a straight line. If that is not the case, the conditions of application are not verified and the test cannot be performed. In this situation, the test will allow you to estimate the coefficient of correlation (the strength of the link) and several tests on the nature of the relationship between them. Another book would be required in order to cover the topic of the linear regression totally, but here are a few important points:

- Linear regression can take multiple forms according to the nature of the variables.
- It is interesting to perform it when the explanatory variables are already known (or supposed) as being relevant for the dependent variable. Otherwise, you might prefer a method that is more explorative and quicker like the chi-squared test.
- The advantage of the linear regression method is that it can quantify the strength of the relationship, but the analysis is asymmetric – you cannot evaluate the dependency of the explanatory variable with the dependent variable at the same time.

Logistic regression

Logistic regression is derived from the linear regression, except the fact that it is focused on the analysis of categorical (non-quantitative) variables, either binomial or multinomial. An **odds ratio** is used to evaluate the probability of a case divided by the probability of a non-case.

The advantage of the method is that, even though the dependent variable needs to be a binomial or multinomial value, the explanatory variables can be continuous (quantitative).

Machine learning

In this section, we will provide an overview of some of the typical methods used inside the field of machine learning that currently finds more and more applications in the analysis of video game players.

Definition

Machine learning was initially a branch of artificial intelligence that was dedicated to the realization of systems that can learn from large data sets. The idea is to assimilate and create rules according to the observed data and then to use them in order to make predictions.

According to these objectives, this field has produced lots of algorithms over the years. These algorithms can now be re-used for other applications, such as data analysis.

We will examine how those algorithms are classified.

Supervised learning

Algorithms derived from supervised learning are supposed to generate a function from labeled training data, where the link between the training example set and the desired output (also called **label**) is made by the analyst himself. From this point and according to the type of algorithms that are used, some parameters usually need to be chosen in order to improve the relevancy of the operation.

In this category, we find methods such as decision tree (which belongs to the hierarchical clustering category), where applications in the video game field can be quite various, and the next paragraph will give an example.

For example, we want to study the path realized by all of the players from their entry into the game to the desired output; for example, buying goods in the game. A decision tree would synthesize the information while displaying the repartition of players at different moments in the game (chosen variables for the algorithms) and would show which path is the most likely to lead to the desired output.

Methods such as decision tree have several advantages:

- They are simple to interpret after a few explanations from the analyst
- They are easily scalable – more variables can easily be included once the best parameters for the algorithms have been identified
- They can also be combined with other decision techniques

On the other hand, if we take the case of a decision tree, the data and variables introduced to perform the algorithm can hardly be categorical, and even with quantitative values, it is important to have variables that are balanced between them.

Unsupervised learning

Unsupervised learning is an explorative way of finding relevant relationships and links inside a large data set of unlabeled variables. The approach is, "we do not know what to search for, but we believe there are some homogeneous groups or specific patterns inside the sample, and we want to find them."

Basically, it is all about finding the hidden structure within a data set. Algorithms derived from unsupervised learning rely on density estimation methods, and the goal is to segment the data in a geographical space, where each dimension of this space corresponds to a given variable. Even though a human cannot imagine a world with more than three dimensions, it is mathematically possible to analyze localizations of a data set in a five-dimensional space or more.

The most common approaches include clustering methods such as K-Means or Mixture of Gaussians. The goal here is to identify homogenous groups of individuals, thanks to the minimization of their standard deviation.

Usually, those algorithms will use computing power to perform a huge amount of iterations (the number of iterations is usually a parameter that can be chosen) on the data set in order to have the best segmentation available.

Once the algorithm has been performed, you will look at the centroids (the mean) of each cluster for each variable in order to pull the relevant information.

As an example for the video game field, you could imagine a data set containing your entire player base with the following variables:

- Number of player versus the player games made
- Number of cooperative games made
- Number of solo games made

The idea here would be to realize a clustering that would highlight the gaming themes and identify groups of players that tend to have the same habits of gaming. Once this is done, you can have a look at the money spent in the game for each group, and see if one of those groups tends to spend a lot of money in the game.

A few important points about the preceding method are as follows:

- As in supervised learning, the choice of variables and their balancing is very important for unsupervised learning. You cannot just introduce asymmetric variables, like money spent in the game and the number of solo games played during the month, because of the mathematical maladjustment of those variables. In this example, the second variable would just totally overwrite the first one because of its numerical weight.
- Remember that those algorithms are made for the exploration – most of the time they will not be perfect, and they constitute an excellent first step towards something more precise, but they can hardly be used to draw conclusions.

Summary

We have provided an overview of some of the typical statistical methods, and what is important to retain is that although there are lots of algorithms and operations available, they usually belong to one of the following categories: description, relationship between variables, and clustering. Their names differ because they are supposed to be applied on different types of variables. In the next chapter, we will see the last part of data analysis, that is about data visualization. Far from trivial, this aspect must not be underestimated because as you will notice, bad data visualization can simply modify the perception of the number and influence your audience.

6

Data Visualization

In this chapter, we will see a very particular field of the data analysis that is the data visualization. Concretely, the topic covers a range of subjects larger than what you may expect at the first sight. According to *Michael Friendly* in *Milestones in the history of thematic cartography, statistical graphics, and data visualization*, published in 2008, data visualization corresponds to the study and optimization of information that has been abstracted in some schematic form, including attributes of variables for the units of information. Globally, the biggest utility of data visualization is certainly to transform complex phenomenon into simple and comprehensible representations.

This chapter will try to list some of the typical data visualization tools and will explain how it is important to mix functionality with aesthetics, without being too sophisticated. As *Fernanda Viegas* and *Martin M. Wattenberg* highlighted it in their article, *How to make data look sexy* (CNN website), an ideal visualization should not only communicate clearly but also stimulate viewer engagement and attention.

First we will study the different forms of representations, and we will have a look at graphic semiology. Secondly, we will describe some of the typical traps in data visualization that you will want to avoid.

Recommendations for good practices

This first section will cover the basics of data visualization as well as some concepts of graphic semiology.

Basic recommendations

A few rules need to be respected while using charts and diagrams. They are as follows:

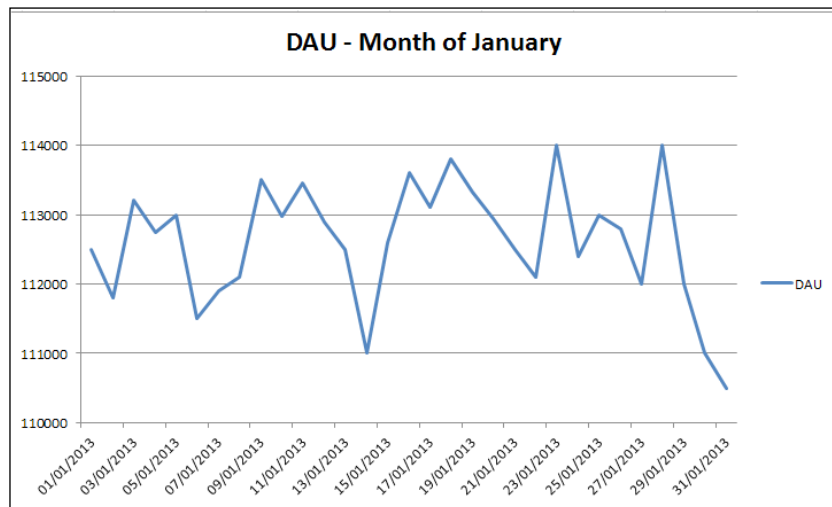
- Never use colors or any significant visual markings if they do not have a link with the data. It may confuse your audience.
- Choose the right type of diagrams for the right data. For example, do not use a line chart for showing separate variables. Generally, a line chart is especially used for showing an evolution over time.
- While showing an evolution over time, choose the right timescale. KPIs usually have many variations of short time period, but those short time evolutions are not usually relevant in order to take any sort of decisions.
- If some data is missing on particular days or for particular periods, do not forget to highlight it with a break on the timeline.
- Be careful with the choices of scale for the variables. Comparison between variables that do not have the same scale can be dangerous. At the same time, be aware of the scale that is chosen for the values of your axis (we will highlight this in the next section).

Typical data visualization tools

This part will present a list of the different charts and diagrams that can be used for data visualization.

Line chart

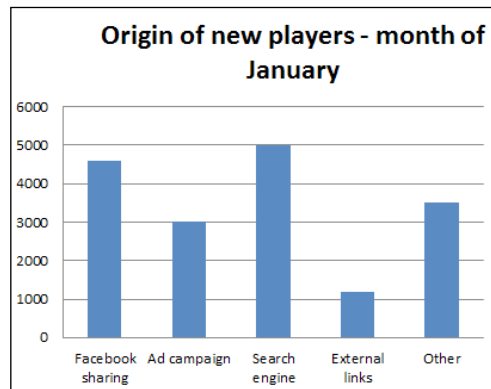
The following screenshot represents a typical line chart:



This sort of graph is one of the most popular and should be used essentially for showing an evolution over time. Considering the video game field, many KPIs (Daily Active User, Average Revenue Per User, and so on) that we have seen in *Chapter 2, Common Key Performance Indicators*, will be displayed on a line chart.

Bar chart

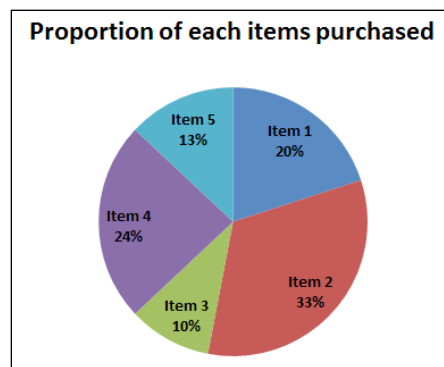
The following screenshot represents a typical bar chart:



This sort of graph is interesting for comparing values between different categories. Referring to video games' KPIs, the origin of your traffic is a good example where a bar chart should be used.

Round chart

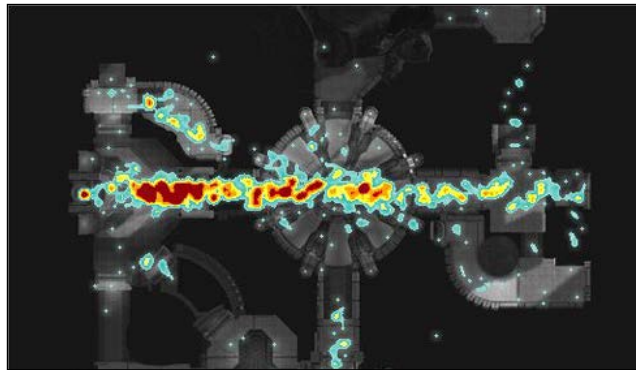
The following screenshot represents a typical round chart:



This sort of graph is designed to be used for visualizing and comparing proportional data (percentage). Despite this restriction, this sort of chart tends to be overused in many situations, but if you do not have a percentage value, you can select a bar chart. Good examples of percentage data could be the proportion of budget that you have allowed to different services or the proportion of each in-app goods bought by all your players.

Heatmap

The following screenshot (<http://www.joystiq.com/2007/11/10/updates-for-bungie-net-halo-3-matchmaking-dlc-on-their-way/>) represents a heatmap from the Halo 3 game, where the color spots represent places where players tend to die more often:



As you can see, this sort of data visualization will be useful when you want to show information from a 3D space if your game contains a geographical dimension.

Now that we have seen some of the most common representations for data visualization, we will highlight some key points about graphic semiology.

Graphic semiology

The graphic semiology field covers all of the rules related to a graphic representation system used for transmitting information (in our case, especially data information). Essentially used for cartography, it is quite useful to know some of its concepts in order to choose from the various ways for displaying data.

One of the pioneers of semiology, *Mr. Jacques Bertin*, distinguishes six different variables that can be adjusted in order to vary the signification of symbols in his book, *Graphic semiology. Diagrams, networks, and maps* (1999).

The first group is composed of the following four variables that are used for differentiating between elements:

- **The orientation:** According to the orientation of your symbol, you can decline different sorts of variables without using a new symbol. It is useful when variables are different from each other in terms of signification.
- **The color:** The color of a shape.
- **The structure:** The sort of the shape (arrows or boxes).
- **The density:** The density of the pattern that is used on a symbol.

The second group is composed of the following variables that are used for displaying the order between values:

- **The size:** One symbol can have different sizes in order to show different quantities of the same element
- **The brightness:** The declination of one color pattern on one symbol can be made through different proportions of white and black

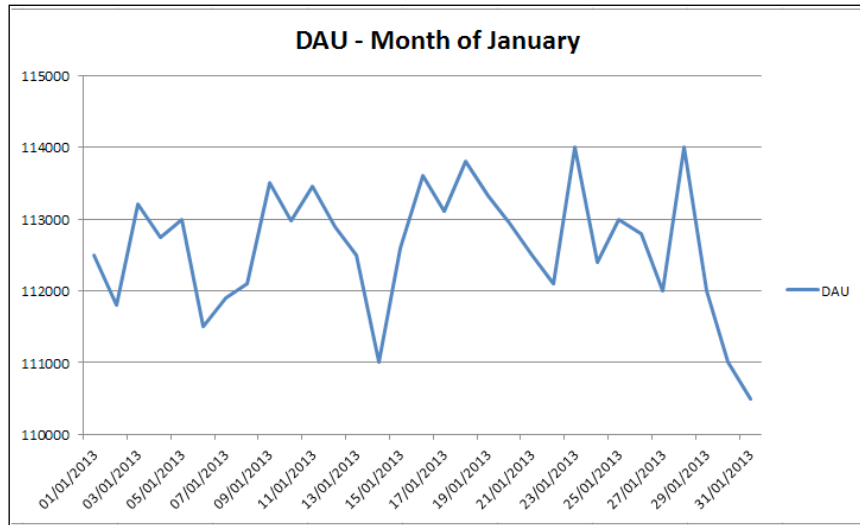
All of these variables are interesting to consider while planning and realizing a map or an outline of tendencies. They can be adjusted differently according to the topics such as business plans, decision trees, and game and level design layouts. The objective is to be as concise and clear as possible.

Typical traps of data visualization

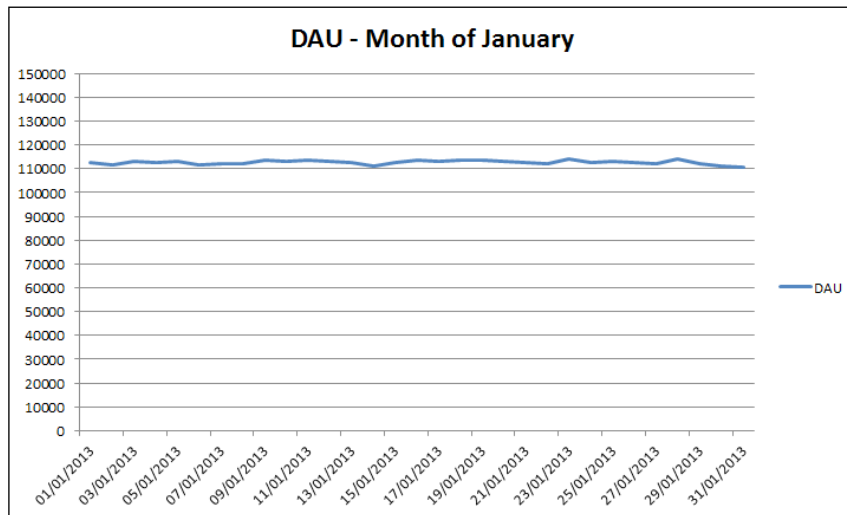
In this section, we will highlight a few example cases where a bad design of the visualization tends to influence the judgment of the reader on the reality of the numbers.

The choice of the scale – values on axis

The following chart displays the evolution of **Daily Active Users (DAU)** on a timeline. Take some time, and look at it carefully:



As we can see, the evolution seems to have lots of variations, and the curve also has a worrying decrease at the end of the month. Now let's have a glimpse at the same chart with 0 as a starting point for the unit on the y axis:



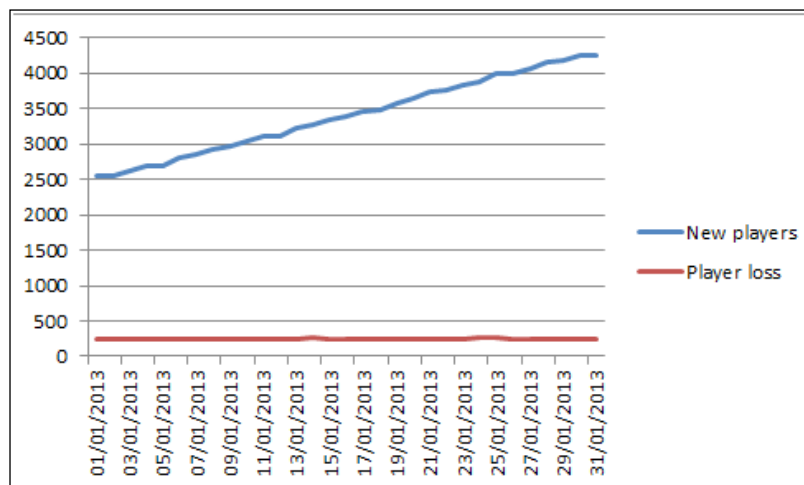
It definitely does not look so frightening! Though the numbers on both charts are identical, only the scale have changed. This example shows how the choice of the scale for your values can considerably change the perception of the numbers. Always think twice before choosing your units for a chart, and try to consider what the best unit for contextualizing your data is.

The previous statement also holds for the ratio between the scale and the size of a shape or a symbol displayed on a chart.

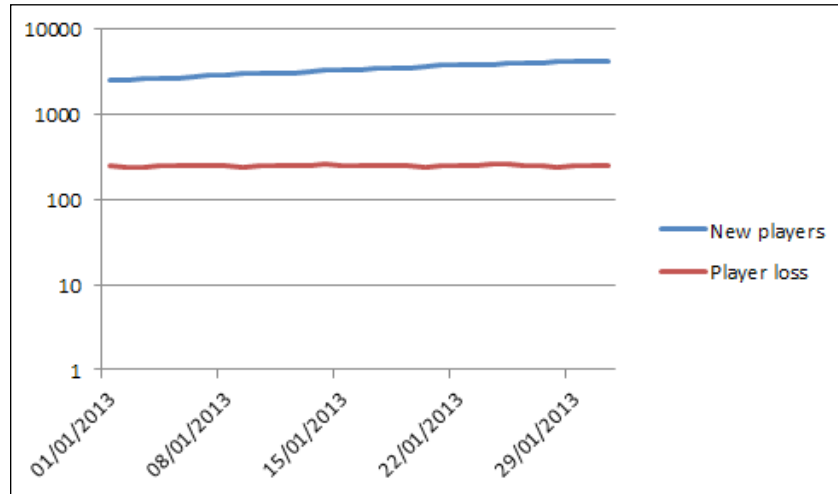
The choice of the scale – equivalence and units between variables

Very simply, it is important to compare what is comparable on a chart. Especially when trying to show correlations or probable links between curves, be very careful about the use of different units because this may lead to wrong interpretations.

An illustration of bad scaling between two variables is shown in the following screenshot:



The preceding screenshot aims at comparing the loss of players during the month (for example, players who did not connect with the game during one month) with the gain of new players. The problem in the first screenshot is that the scale chosen on the y axis tends to promote the tilt of the curve of new players gained because the starting values for both variables are not the same. A way to solve this problem is to use a logarithmic scale, as shown in the following screenshot:



As we can see, the evolution of both variables on the timeline is now represented properly. To be more specific, each variable is now represented according to its numerical weight.

Summary

This chapter on data visualization has now come to an end, and you should be able to avoid most of the basic errors related to it. As a data analyst, it is important to take care of this aspect because though numbers and charts seem more obvious to you, you must not forget that this is not the case for managers or producers. There is no doubt that a few errors in data visualization can drastically change the quality of your presentations.

In the next chapter, we will see the limits of game analytics and how this new tool should be handled in general.

7

Limits of Game Data Analysis

In this last chapter, we will see some general recommendations about game analytics and the kinds of limits they have. The first section will detail the philosophy and objectives that should guide the use of game analytics, and the second section will outline their limits.

Which game analytics should be used

This section will focus on the role data should take in your production process. As a studio, the first step is to identify your needs and to choose the goals you will attribute to game analytics.

Game analytics as a tool

Firstly, it is important to understand that game analytics are a tool, which means they can serve several purposes. You can use them for marketing, science, sociological studies, and so on. Following this statement, you will need different tools and different approaches to reach your goal. As this book has tried to highlight it, tools are chosen according to problems, regardless if the choice is technique or analysis. You must not choose a tool because it is said to be the best performing tool ever made, or because it is fashionable. Instead, you must choose a tool because it is said to be the most efficient tool for your needs. Try to answer the following questions:

- What are the long-term uses I plan to do with game analytics? Is it simply reporting the Key Performance Indicators or is it the building a user-centric framework for deep analysis?
- What are the types and the level of skills of the people who will work on it? Do I have all of the skills, from data scientists to game analysts, or do I need to choose a solution which will offset some lacks in a particular field?

- How much data will be collected? How do I plan to deal with possible peaks of frequentation?
- How do I adapt temporalities of reporting and analysis with the rhythm of production I have on my project? Do I split them weekly or monthly?
- What are the main goals of my process? Do I want to build a predictive model (for example, based on correlations) in order to define the next acquisition campaign I will run? Do I want to increase the monetization rate on the current player base? Do I want to perform A/B testing? And the list goes on.

Game analytics must serve your team

Secondly, it is important to ensure that the use of game analytics must serve your team as a whole. They should not have any disagreements about the long-term objectives that you have chosen.

They must accompany it and especially improve it, but the general objective should remain the same. Given the current state of the field, withdrawing the "human touch" from the design process entirely and listening only to data would be a mistake. That's why the game analytics process should be thought through the prism of your own team; and therefore, should be presented as a new tool. This will help them to make good decisions for the game.

The best example for the democratization of "game analytics way of thinking" inside your team is certainly the A/B testing aspect. If you experience debates about particular features in the game, instead of taking part you can propose to use A/B tests for some of those features.

Following this, there are no particular limits to the use of the tool. A game designer can test different balancing on the virtual economy of a game and an artist can experience different graphic styles.

When starting, focus your attention on simple practices

If you are new to the field, the following list may help you to start defining your first objectives. It contains most of the typical use for online games, especially free-to-play games:

- Producing KPIs on a weekly or monthly basis, according to your needs. These KPIs will help you to orient the upcoming development of your game and to anticipate the return on investment of your acquisition campaigns.

- Identifying if some of the steps of your tutorial phase are poorly designed; for example, if you have a sudden player loss at a particular step of your tutorial.
- On the same idea, having the loss of players at each level is also very useful to improve the general balancing of your game, especially the progress curve and the difficulty. This topic is more important if you have a part of your business model based on purchasable goods, which can increase the progression rate of the player.
- You can evaluate which area and which purchasable goods of your game are generating the best income.
- You can perform A/B testing on particular key features of your game in order to see which ones are the most efficient.

What game analytics should not be used for

On the other hand, there are a few limits that you need to know before using methods and processes from game analytics.

Keep away from numbers

You must always be careful about the fact that numbers are used to represent a given situation during a "T" instant. From this statement, the predictive models must always be revised and improved. they should never be considered as the perfect truth.

In order for the process to be efficient, it is quite important to keep research on the data inside the structure defined by the initial goals. Otherwise, you might split your efforts and no actionable insights would be identified.

In other words, numbers must remain at their place. They are a tool in the hands of a human subject, and they should not become an obsession. Try to reason if they make any sense and if you are asking the right question.

Practices that need to be avoided

As mentioned in the the previous section, if you are new to this field, be aware of the following situations:

- Data cannot dictate the full content of your next update. If it is the case, you may first re-evaluate the general intention behind your product and talk with the game designer.

- When starting, try to avoid complex questions that involve external factors in the game, even if they seem crucial for you. For example, trying to understand why people stopped playing your game over a long period of time is usually impossible. Old players might stop playing because another game came out or they just got bored. Data cannot make miracles at this point of the engagement.
- Data must not take too much amplex in the creative process. There are some human intentions and ideas, and only then the data comes in order to verify and improve the potential success of those intentions.
- Data must not slow down the performances of the game. One of the common methods to avoid this is to send the data when the player logs in or logs out and not at each click or each action.

Summary

This is the end of this chapter, and the most important thing you need to remember about game analytics in general is the importance of the definition of your objectives. The reason why you choose this tool instead of another (and this book has tried to list a maximum of them, from data mining to pure analysis) is because it fits your needs as much as possible. This statement is true at every stage of the reflection process which surrounds game analytics, from the choice of the storage solution to the type of analysis you want to perform.

The rising of a fully-connected state in the video game industry offers developers the opportunity to change the way they create games, but there is no doubt that the level of maturation related to this tool has not reached its maximum yet. Therefore, even if the benefits of game analytics are great, be prepared to make mistakes as well; and keep your own process open to various criticisms from your team.

Index

A

A/B testing 37, 38
acquisition KPIs
 Cost per acquisition 22
 K-Factor 22
 New players 21
 Source of traffic 22
 Tutorial loss 22
 Virality 22
attrition rate 23

B

bell curve 48
Buy button 40

C

central point
 about 46
 highlighting, indicators 46
central point indicators
 arithmetic mean 46
 median 46
chi-squared test
 about 50
 characteristics 51
 using 50
commercial solutions
 Flurry Analytics 33
 Honey Tracks 32
 Kontagent 32
correlation 49

D

Daily Active Users (DAU) 62
data analysis
 commercial solutions 32, 33
 development environments 27- 29
data analysis tools
 about 33
 commercial tools 34
 open source tools 33
data mining tools
 about 29, 30
 Facebook Insights 30, 31
 Google Analytics 31
data visualization
 about 57
 recommendations, for good practices 57, 58
 tools 58-60
 traps 61-64
data visualization tools
 bar chart 59
 heatmap 60
 line chart 58, 59
 round chart 59, 60
data visualization traps
 scale, choosing on charts 62-64
decision tree
 advantages 53
 disadvantages 53
desire for challenge
 about 10
 game challenges 11
 player challenges 11

desire for imagination

- about 12
- desire for discovery 12
- desire for distraction 14
- desire for emotions 13
- desire for entertainment 14
- desire for immersion 14
- desire for romping 15
- desire for sensations 13
- storyline 14

desire for reward

- desire for achievement 10
- desire for collection 10
- desire for ownership 8, 9
- desire for reputation 9

development environments for data

analysis

- Hadoop 29
- Hive 29
- MySQL 27, 28
- NoSQL 28, 29

DGD1 model 7

dispersion indicators

- interquartile 47
- standard deviation 47
- variance 47

F

Facebook Insights 30, 31

Farmville game 6

flame event 13

Flurry Analytics solution 33

free-to-play model 5

friction 41

G

game acquisition quality

- classifying 23

game analytics

- about 6, 7
- objectives, defining 66, 67
- purpose 7
- situations, avoiding 67, 68

testing 37-39

using, limits 67, 68

using, objectives 65, 66

game analytics objectives

defining 66, 67

game analytics, using as efficient tool 65, 66, 67

game analytics, testing

A/B testing 37, 38

good practice recommendations 38, 39

multivariate testing 38

game engagement

about 15

free-to-play model 16

saleable items typology 16, 17

video game, using as service 15

game themes

about 7, 8

desire for challenge 10, 11

desire for imagination 12-15

desire for reward 8, 9, 10

references 17

Gephi 34

Google Analytics 31

graphic semiology

about 60

variables, used for elements

distinguishing 61

graphic semiology variables

brightness 61

color 61

density 61

orientation 61

size 61

structure 61

H

Hadoop

about 29

Hive, adding onto 29

Hadoop Distributed File System (HDFS) 29

Hive framework

adding, onto Hadoop 29

Honey Tracks solution 32

I

Insights Dashboard 30

K

Key Performance Indicator. *See* KPI

Kontagent solution 32

KPI

- about 19, 47
- selecting, criteria 20
- working with 25, 26

KPI framework

- acquisition 21, 22
- monetization 24, 25
- retention 23, 24
- video game KPIs, organizing 20, 21

L

label 53

linear regression

- about 51, 52
- characteristics 52

logistic regression

- about 52
- advantage 52

M

machine learning

- about 52
- supervised learning 53
- unsupervised learning 54, 55

MapReduce system

 29

Massive Multiplayer Online Role-Playing Game (MMORPG) 9

monetization KPIs

- Average revenue per paying users 24
- Average revenue per users 24
- Conversion rate 25
- Items gain 25
- Lifetime value 24
- Paying users 24
- Source of income 25

monetization pop up

- case study 40
- feature inventory 40-42
- versioning, examples 42-44

multivariate testing 38

N

NotOnlySQL (NoSQL) 28

O

odd ratio 52

Open Graph Insights

- about 30
- metrics 30, 31

Open Graph Insights metrics

- demographics 31
- site engagement 30
- story clicks 31
- story comments 31
- story impressions 31
- story likes 31
- timeline unit activity 31

open source tools

- Gephi 34
- Rapid Miner 35
- R-Project 34
- SPSS 34
- Statistica 35
- Tableau Desktop 36

P

page-centric tool 31

prototype-and-iterate method 6

R

Rapid Miner

- about 35

regression

 49

retention KPIs

- Average gaming time 24
- Average number of game sessions 24

Churn rate 23
Daily active user 23
Monthly active user 23
Retention curve 24
Return Of Investment (ROI) 33
R-Project 34

S

saleable items typology
desire for collection 16
desire for competition 16
desire for complexity and difficulty 16
desire for discovery 16
desire for distraction 17
desire for emotions 16
desire for immersion 17
desire for ownership 16
desire for reputation 16
desire for romping 17
SPSS 34
SQL 27
Statistica 35
statistical distribution
about 48
Poisson distribution 49
student distribution 48
statistical laws 48
statistical methods
central tendencies 46, 47
chi-squared test 50, 51
describing 45
dispersion tendencies 47

linear regression 51, 52
logistic regression 52
statistical distribution laws 48
variables correlation 49
variables regression 49
variable types 49, 50
Statistical Package for the Social Sciences.
See SPSS

Structured Query Language. *See SQL*

supervised learning
about 53
methods, decision tree 53

T

Tableau Desktop 36

U

unsupervised learning 54, 55

V

variables
correlation between 49
regression between 49
types 49, 50
variable types
binomial variable 50
continuous variable 50
discrete variable 49
ordinal qualitative variable 50
qualitative variable 50



Thank you for buying **Game Data Analysis – Tools and Methods**

About Packt Publishing

Packt, pronounced 'packed', published its first book "*Mastering phpMyAdmin for Effective MySQL Management*" in April 2004 and subsequently continued to specialize in publishing highly focused books on specific technologies and solutions.

Our books and publications share the experiences of your fellow IT professionals in adapting and customizing today's systems, applications, and frameworks. Our solution based books give you the knowledge and power to customize the software and technologies you're using to get the job done. Packt books are more specific and less general than the IT books you have seen in the past. Our unique business model allows us to bring you more focused information, giving you more of what you need to know, and less of what you don't.

Packt is a modern, yet unique publishing company, which focuses on producing quality, cutting-edge books for communities of developers, administrators, and newbies alike. For more information, please visit our website: www.packtpub.com.

Writing for Packt

We welcome all inquiries from people who are interested in authoring. Book proposals should be sent to author@packtpub.com. If your book idea is still at an early stage and you would like to discuss it first before writing a formal book proposal, contact us; one of our commissioning editors will get in touch with you.

We're not just looking for published authors; if you have strong technical skills but no writing experience, our experienced editors can help you develop a writing career, or simply get some additional reward for your expertise.

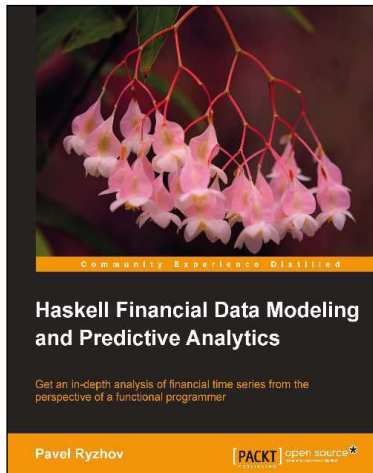


Piwik Web Analytics Essentials

ISBN: 9-781-84951-848-2 Paperback: 326 pages

A complete guide to tracking visitors on your websites, e-commerce shopping carts, and apps using Piwik Web Analytics

1. Track customers on your websites and apps and improve conversion rates right away
2. Customize your tracking for every application and situation
3. Embed Piwik's tracking capabilities in other applications



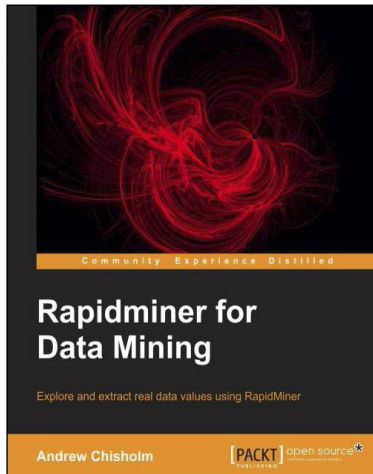
Haskell Financial Data Modeling and Predictive Analytics

ISBN: 9-781-78216-943-7 Paperback: 112 pages

Get an in-depth analysis of financial time series from the perspective of a functional programmer

1. Understand the foundations of financial stochastic processes
2. Build robust models quickly and efficiently
3. Tackle the complexity of parallel programming

Please check www.PacktPub.com for information on our titles

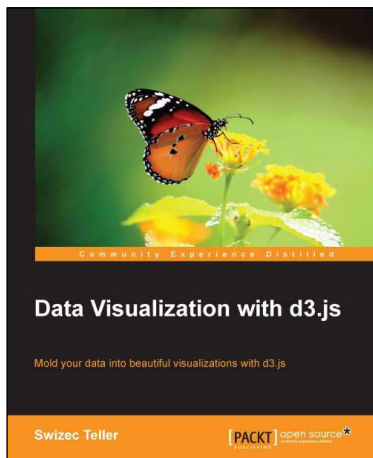


Rapidminer for Data Mining

ISBN: 9-781-78216-933-8 Paperback: 150 pages

Explore and extract real data values using RapidMiner

1. See how to import, parse, and structure your data quickly and effectively
2. Understand the visualization possibilities and be inspired to use these with your own data
3. Structured in a modular way to adhere to standard industry processes



Data Visualization with d3.js

ISBN: 9-781-78216-000-7 Paperback: 194 pages

Mold your data into beautiful visualizations with d3.js

1. Build blocks of web visualizations
2. Learn visualization with detailed walkthroughs
3. Learn to use data more effectively
4. Animate with d3.js
5. Design good visualizations

Please check www.PacktPub.com for information on our titles