Data Science Assignment 1: Visual analytics

Diego Sourbag (s3699420), Reinier Bos (s3703177)

March 13, 2025

1 Introduction

In this report, we will describe how to use the dashboard for the Emerald IT software business. They created the app *Complete Reference for Dungeons and Dragons 5*, which is an app used to play Dungeons and Dragons. The company provided us with data on the usage of the app. This is data about the sales, crashes and reviews of the app. The data covers the months June until December of 2022. This data is processed, cleaned, and visualized in a web-based dashboard.

There are multiple things that can be done with the dashboard. First of all, it gives a quick and easy overview of the data. The different widgets show different overviews. The widgets have tabs to filter on specific attributes. You can hover over the graphs to get detailed information on the graphs. This is an overview of the different graphs in the dashboard:

Component Name	**Functionality**
Monthly Sales & Transactions Chart	Displays sales amounts and transaction counts per month with dual y-axes for comparison.
Sales by SKU Chart	Shows sales performance based on SKU attributes, distinguishing between premium and unlock character manager sales.
Top 15 Countries by Sales Chart	Highlights the sales amount per country using a horizontal bar chart.
Sales by Day of the Week Chart	Visualizes sales distribution across weekdays to identify trends in purchasing behavior.
Sales by Hour of the Day Chart	Represents the hourly distribution of sales to detect peak transaction periods.
Daily Crashes Over Time Chart	Tracks the number of application crashes per day to analyze stability trends.
Correlation Between Crashes and Ratings Chart	Illustrates the correlation between app crashes and user ratings.
Sales Heatmap	Displays a geographical representation of sales by region or country.
Rating Heatmap	Visualizes average rating across different locations.
Sales & Transactions Table	Provides a detailed tabular view of total sales data, and average rating

Table 1: Dashboard Components and Their Functionalities

2 Visualization

2.1 Sales Volume

The first visualization shows the sales and transaction volume over time. The amounts of sales and transactions are displayed on one graph, where the amounts of sales are visible on the left y-axis, and the amounts of transactions are shown on the right y-axis. The months are displayed on the x-axis. The sales data are grouped by month, while taking the sum of all charged amounts and counting the order numbers. We chose to show this on a line chart to show the sales volume over time.

2.2 Attributes and filtering

The second widget consists of multiple graphs. They all show different attributes or aspects of the sales volume. The first graph shows the division of total sales over the different attributes, namely premium and unlocking characters. For each of these attributes, the sum of all sales is calculated per month. This is shown in a line chart, to see the development over time.

Then there are three graphs displaying the sales per country, per day of the week, and per hour of the day. For this graph, we chose to visualize it in histograms, as it easily displays the difference between categorical variables. The sales per country are calculated by grouping the data in a country and calculating the sum of the sales. Only the first 10 countries are shown, as they are the most relevant. The sales per day of the week and hour of the day are calculated by converting the date of the sale to the day of the week or hour of the day, and grouping the sales accordingly.

2.3 Ratings vs Stability

The third visualization examines the relationship between app ratings and system stability. First of all, the crashes are displayed over time. This gives a quick view of the stability of the app. The second graph is a scatterplot of the daily crashes and reviews with a trend line. Here you can see the correlation between a higher amount of crashes, and lower ratings. The third graph shows the average rating and amount of crashes per month. The two variables are displayed over time on both y-axis. The last graph show the stability score, which is a Key Performance Indicator. Stability is calculated by:

$$\text{Stability Score} = \frac{\text{Daily Average Rating}}{1 + \text{Daily Crashes}}$$

The stability score is equal to the Daily Average Rating when there are 0 crashes. If there are crashes, the score will be lower. This gives a quick overview of the stability of the app.

2.4 Geographical Development

The last widget shows the division of sales volume and ratings per country on a map. This gives an easy overview of how it is divided into different countries. You can hover over the map to see country-specific data. The map is created by converting the two letter country code to a three letter country code. The data is grouped by country, while adding the sales volume and calculating the average rating. With the Geopandas package, this is displayed on a world map, so that it can be easily compared.

3 Decision making

The company is prepared to launch a marketing campaign next year. Therefore, it is essential to select the right target markets to maximize growth. To ensure a data-driven approach, emerging countries are targeted. Emerging countries are defined by a consistent increase in sales and transactions with a positive customer rating. The Sales per Country visualization highlights the top 15 markets by total sales. The top 3 countries include the United States, Great Britain, and Canada. Although these countries account for a large share of total sales volume, they are not considered emerging countries. These markets are already saturated. Instead, our focus is on countries where sales and transactions are rapidly growing. Looking at table, some of the medium-sized countries based on sales are visible, such as the Netherlands, Mexico, and Australia. By hovering above these countries in the Sales Map, it is noticeable that of these three countries have more than doubled the amount of sales in December compared to June. This significant upward trend in sales volume stands out. To verify the positive expectation of these countries, checking with the Global App Ratings map is an option. By checking if the ratings remain stable or increase during the growth in sales, emerging countries can be determined. The average ratings for the three countries have remained stable. Therefore, the Netherlands, Mexico, and Australia are suggested as a target group for the marketing campaign next year.

4 Satisfied and Critical Customers

To determine the ratings of the countries, see the table below on the left. By clicking on the header of the Average Rating column 2 times, the rating is sorted from lowest to highest rating. Therefore, it is easy to see which countries relatively rated the app the most negatively. It is a tie, because both Paraguay and Botswana have an average rating of one. This makes these two countries relatively have the most users to rate the app negatively. Using the same table, it is easy to see which countries have the highest rating. There are multiple counties with the highest average rating of five, such as Belarus and Bulgaria. The five-star-rated country with the highest sales volume is Switzerland. During the five months, there was a release with a serious bug that affected all Android 7 users. Looking at the crashes per day it should be between the tenth and twentieth of November, as this period is followed by a new all-time high within the line chart for crashes, from that date onward. In the graph Monthly Average Rating and Total Crashes it is clearly visible that in the last 2 months the total crashes have skyrocketed.