UC3DVS102 - Data Visualization final assessment

Simen Aronsen

ABSTRACT

The purpose of this task is to make an interactive dashboard of one dataset with information on football players profiles. The main task for this dataset is to visualize some aspects of the players and also make a dashboard that is interactive so the users/stakeholders can easily access the data on a web page and interact with the dashboard to get the desired information. The second task is to make a static visualization chart deriving insight into a dataset containing Amazon's Top 50 bestselling books from 2009 to 2019. The purpose of this task is to get some insight from the data that can be presented to customers/stakeholders that may find it interesting.

1 INTRODUCTION

For the first task, we will use Tableau for the visualization of the data. This software has been proven to be the best one for usage when it comes to creating interactive dashboards that can easily be published on a web page so it can be shared with multiple stakeholders. For that specific task, we want to present some information from the data that the user interacts with. The user can click on the different charts to view some specific information from the dataset. For the second task, we will make a static visualization that shows the data for the amazon books dataset. For this specific task, we will be using Power BI to Visualize the data in question.

2 DATA GATHERING AND PREPARATION

2.1 Footballers Profile Dashboard

First, we download the dataset from the course page moodle. The dataset itself is gathered from Kaggle.com and this site is entered to study how the data is organized and but together. After this, the data is imported into a Jupyter notebook and the pandas library is used to explore the data to see if it needs data transformation. It is decided to not transform any data at this point in the research so the data is imported into the chosen platform Tableau to make the visualization.

2.2 Amazon Bookshelf

First, we download the dataset from the course page moodle. The dataset itself is gathered from Kaggle.com and this site is entered to study how the data is organized and but together. After this, the data is imported into a jupyter notebook and the pandas library is used to explore the data to see if it needs data transformation. It is decided to not transform any data at this point in the research so the data is imported into the chosen platform Power BI to make the visualization.

3 DESIGN AND IMPLEMENTATION

3.1 Footballers Profile Dashboard

The design choice for this task was to make an interactive dashboard with four different charts that the user could interact with on a web page. The first one is a Scatterplot that displays the preferred foot used by players using the average dribbling and average ball

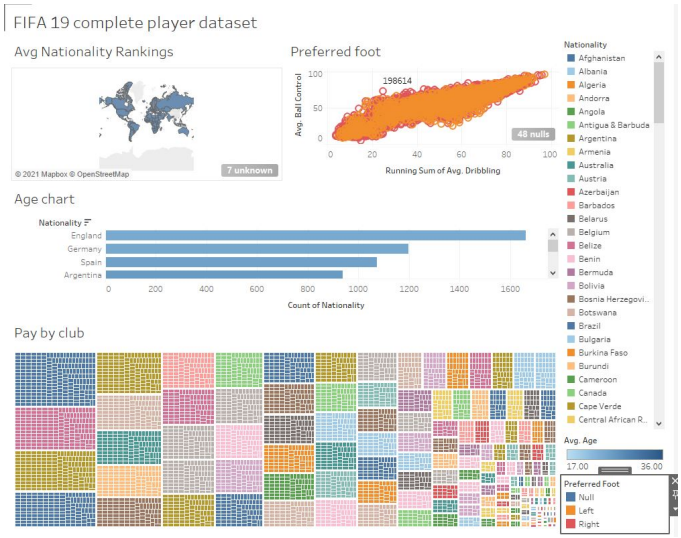

Figure 1: Dashboard topic 1

control variables. The player ID is used to identify the player and the user can click or hover over the player to see the preferred foot, ID, avg. ball control and running sum of avg. dribbling. The second chart is showing the Pay by Club where information about the Club, Nationality and maximum wage is displayed on cards for all the clubs. This chart also makes it easy to see where most of the clubs in the dataset are localized. The third chart is displaying the avg. Nationality Rankings on a map that the user can hover or click on the specific country to see the avg. overall and penalties score. The last chart is a bar chart that shows the avg. age of the players in regards to nationality and a count of the players with this age. All of these charts are implemented into a Dashboard with the charts in the middle for the user to interact with and a Legend on the right side with information on the colour codes that are used for the different charts. The user is also able to click in the different countries to display the information in interest. Figure 1 shows an image of this dashboard. URL of the dashboard: https://dub01.online.tableau.com/#/site/fifa19completeplayerdataset/workbooks/721323?:origin=card_share_link

3.2 Amazon Bookshelf

For this task, the main objective was to show some basic information extracted from the Amazon books dataset. It was decided to use Power BI to visualize the reviews by year and genre. This is a simple bar chart that is clustered with the genres, side by side, separated by year on the x-axis and the number of reviews on the y-axis. This information can showcase the different interests and engagement of the users of amazon in particular years to the different genres and can help future customers choose books from a specific time period in regards to genre and interests. The design are simplistic

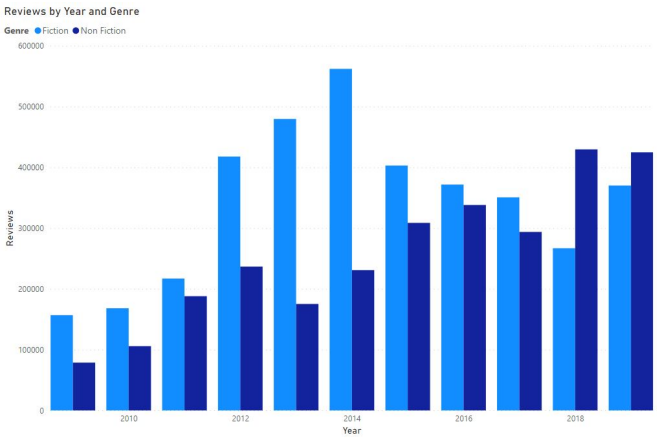

Figure 2: Static chart topic 2

with light blue and dark blue on the columns to easily separate the two genres in question. Figure 2 shows the design for the task.

4 CONCLUSION

Two different topics were discussed in this paper. An interactive dashboard visualization was created using the Tableau software and a static visualization chart using Power BI software. Both tasks used different software to visualize the data and the reasoning for this implementation choice was to explore both of the software to learn from them and also because of limitations in publishing abilities. For future projects in visualization, it could be beneficial to look at an implementation using other tools like the python libraries Plotly and Matplotlib for data transformation and visualization. The scope for the task was chosen from the background knowledge of the data in question and also the knowledge of the chosen software. The desired results were successfully achieved by the author of this paper. Improvements could be implemented in the design to make it more interactive and derive more insights from the data.