
DATA SOURCES
ACTION
MACHINE LEARNING
STATISTICAL MODELS
REAL-TIME PROCESSING
REAL-TIME DASHBOARDS
NOTIFICATIONS
DATA SHEET
Portfolio Projects
Loredana Mot Pribac

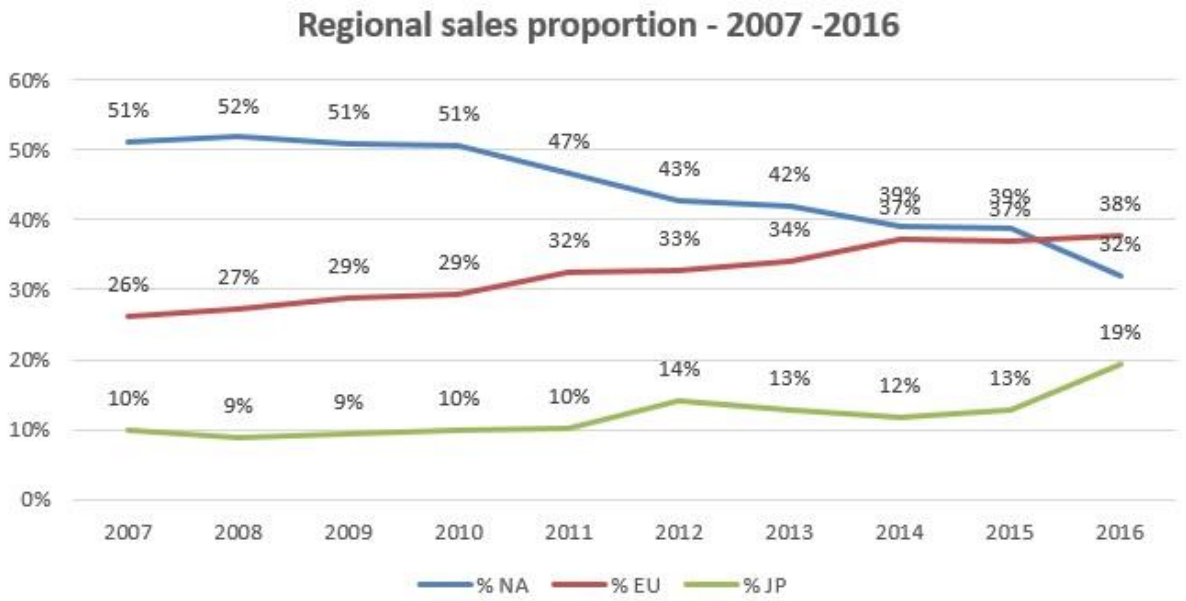
Contents

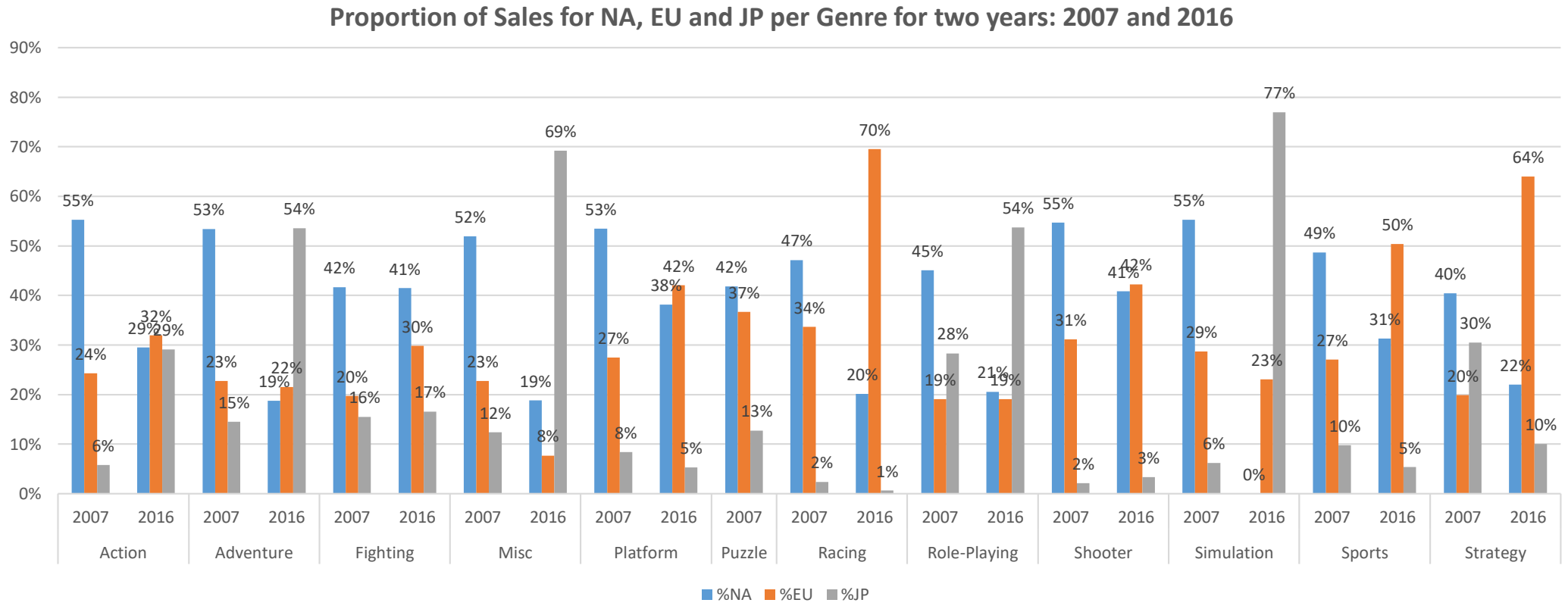
P1.Analyzing global video game sales

- Sales Forecast for GameCo. The company wants answers that will help make a decision regarding marketing budget for each zone.
- The data used for analysis is from GameCo.

- I used the tools I had worked with over the course to derive new insights. It’s important marketing teams know which genres are most popular in each region as well as sales in each region.
- I want that the way I choose to do the summaries, groupings, and visualizations, to make me lead to insights of this data set. I want my project to give some answers to the GameCo.
- I selected the data for my project for 10 years’ period: 2007 – 2016, and I started with the Proportion of North America Sales, European Union Sales and Japan Sales reported to Global Sales.

Relevant graphic representations:





Source: My own realization

Conclusions and recommendation

- North America was the biggest seller for video game from 2007 until 2015, but now European Union sales are bigger than North America.
- European Union the market shows a regular small growth trend every year.
- Japan shows an increase in the last 10 years.
- **Main insight: So the trend for these three regions is the same, even when we talk about global sales for video games, even when we talk about sales on genres or platforms.**
- **Two reason for the trend of the sales for each region: every year sales, for every region depend on the releases of new video games of the biggest companies and the amount of money they spend on marketing in every region.**

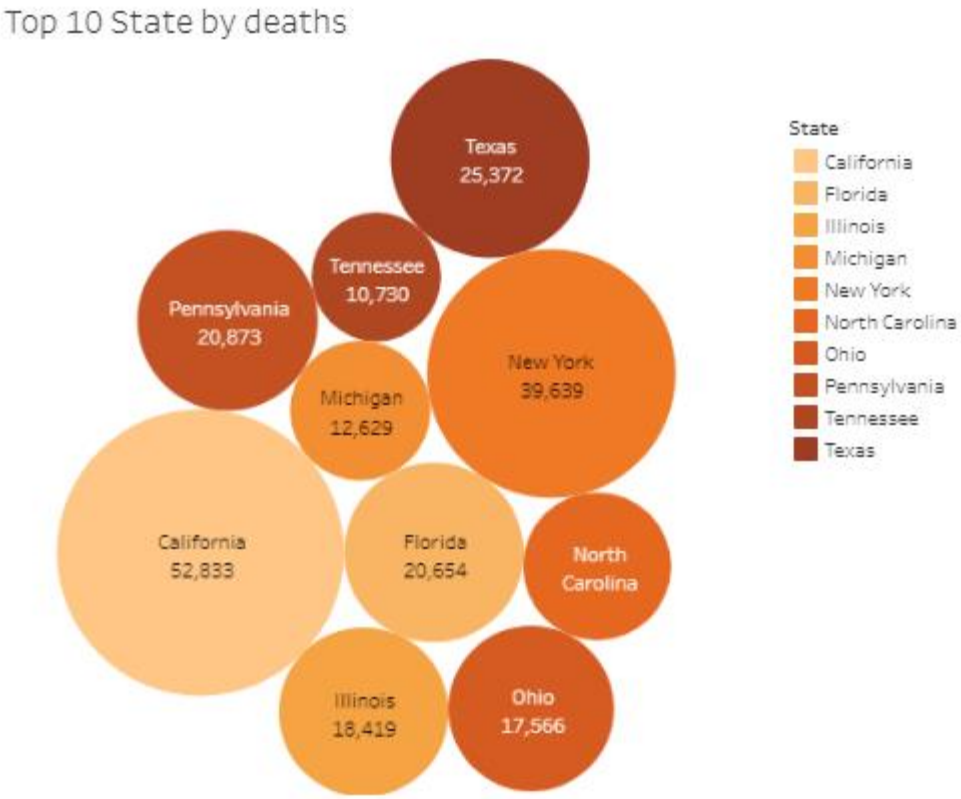
- **Recommendation: In terms of marketing spend the investments in Japan and European Union should increase and the investment in North America should decrease.**

P2.Preparing for flu season in the U.S.

- **Goal:** To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.
- **Motivation:** The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.
- **Objective:** Determine when to send staff, and how many, to each state.
- **Scope:** The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.
- The databases used for the analysis are: US Census Bureau, CDC and questionnaires.

Relevant graphic representations:

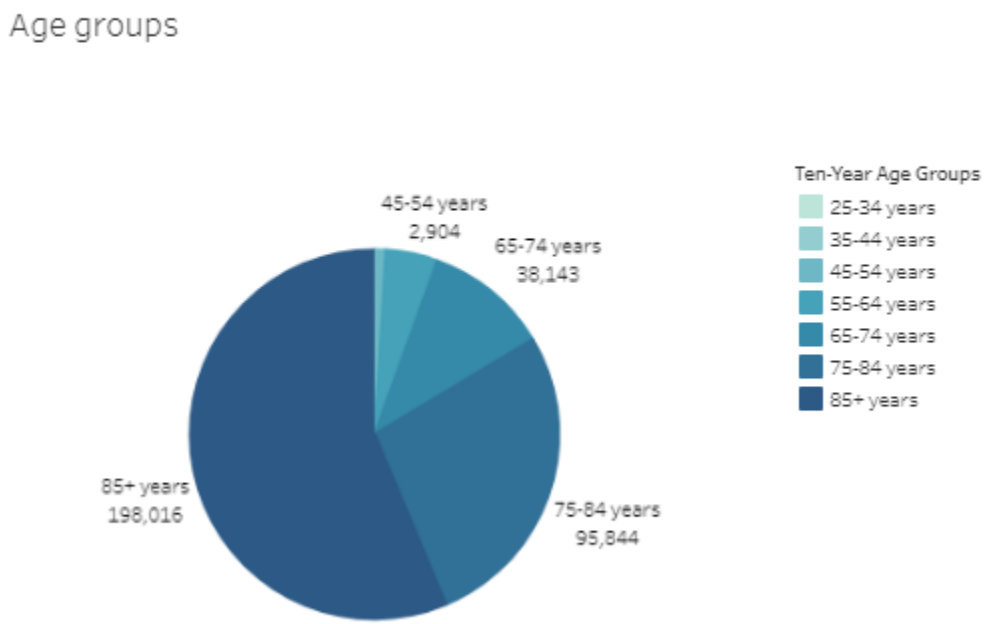
The states with big population are the most affected.

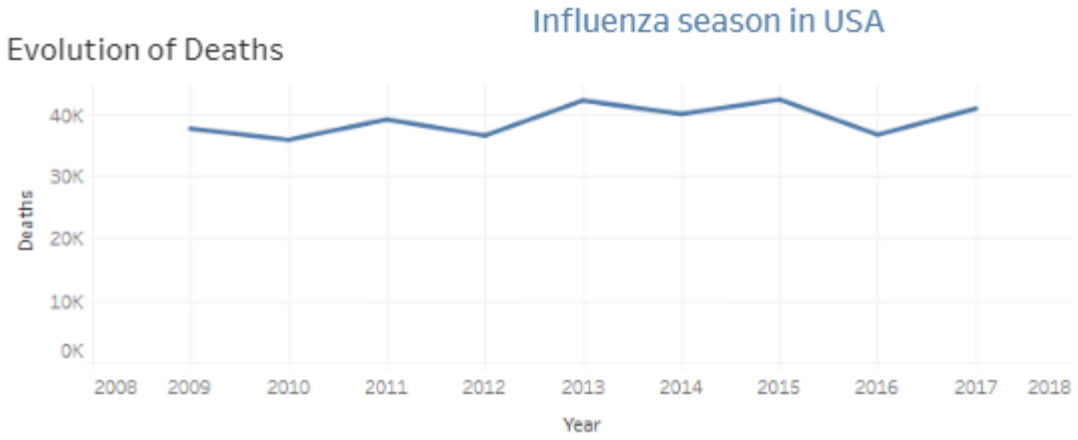


Source: My own realization in Tableau

Relevant graphic representations:

The Age groups show us that the most affected are: 85+ years, at big distance are 75-84 years, and the next age group is 65-74 years. Like it is to expect.





Source: My own realization in Tableau

Conclusions and data limitations

- **The most affected by the flu are old persons.**
- **The countries with the biggest population are the most affected by flu every year.**
- **Data limitations:**
- If I had a dataset with number of staff for each state, I would have been able to do a complete analysis.
- Some state for some variables had missing values. So the resultants of the project have this minus.
- Create a dataset that has the information about all the changes in activities that staff is doing. So we can monitor the impact of the staffing changes.
- We can try to have in staff dataset this: staff for each state in every month of the year, number of staff that was sent in other states to help, a number of sick staff in the wintertime, a list with hospitals and clinics with lack of personal.
- All the visualizations are here, in Tableau:
- https://public.tableau.com/profile/loredana.mot.pribac#!/vizhome/InfluenzainUSA/InfluenzaseasoninUSA?publish=yes

P3. Answering business questions for an online video rental company: Rockbuster Stealth Data Analysis Project

- The Rockbuster Stealth Management Board has asked a series of business questions and they expect data-driven answers that they can use for their 2020 company strategy.
- Here are the main questions they'd like to answer:

● Which countries are Rockbuster customers based in?

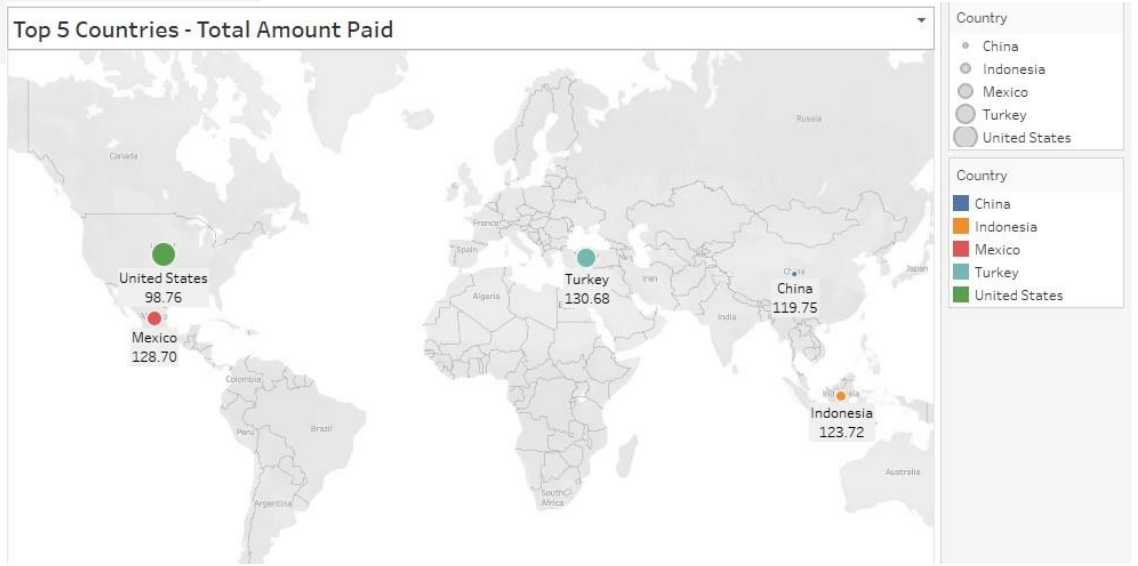
● Where are customers with a high lifetime value-based?

● Do sales figures vary between geographic regions?

The data set that contains information about Rockbusters film inventory, customers, and payments, among other things. All this information is collected by them.

P3.Answering business questions for an online video rental company: Rockbuster Stealth Data Analysis Project

- To answer the questions posed by the different departments, I did query the data using SQL.
- **After I did cleaning data, ordering, limiting, filtering, grouping data, joining tables of data and subqueries, I did this:**
- Descriptive Statistics
- Top 10 Countries – Number of customers
- Top 5 Countries – Total amount paid



Relevant graphic representations:

Source: My own realization in Tableau

Conclusions and recommendations

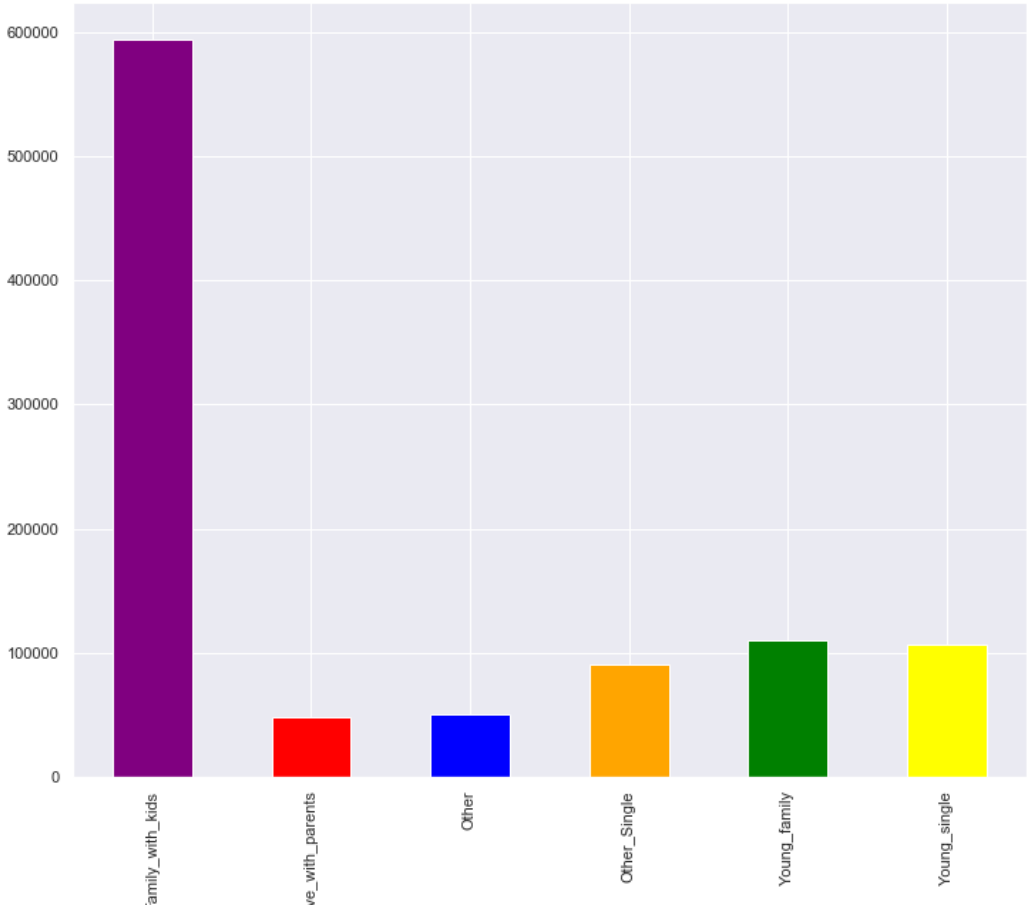
- I find that AVG for rental rate is 3 $, the AVG for replacement cost is 30 $ and AVG for the rental duration is 5 days.
- The most movies have ratings: PG-13.
- The Store 1 is used often.
- The biggest number of customers has this country: India, China, and the US. This country has a lot of the population, so to be in the top 3 is good for the company.
- The countries with the total amount paid are Turkey, Mexico, and Indonesia.
- I analyzed the data from the Rockbuster dataset.
- I used SQL, Excel, and Tableau.
- The results of the dataset analyzed are important for the decisions of Rockbuster business managers.
- **They must invest the money for marketing in countries that have a big population and the countries are not economically bankrupt, so the people can have money to spend on this.**

P4.Marketing strategy for an online grocery store: Instacart Grocery Basket Analysis

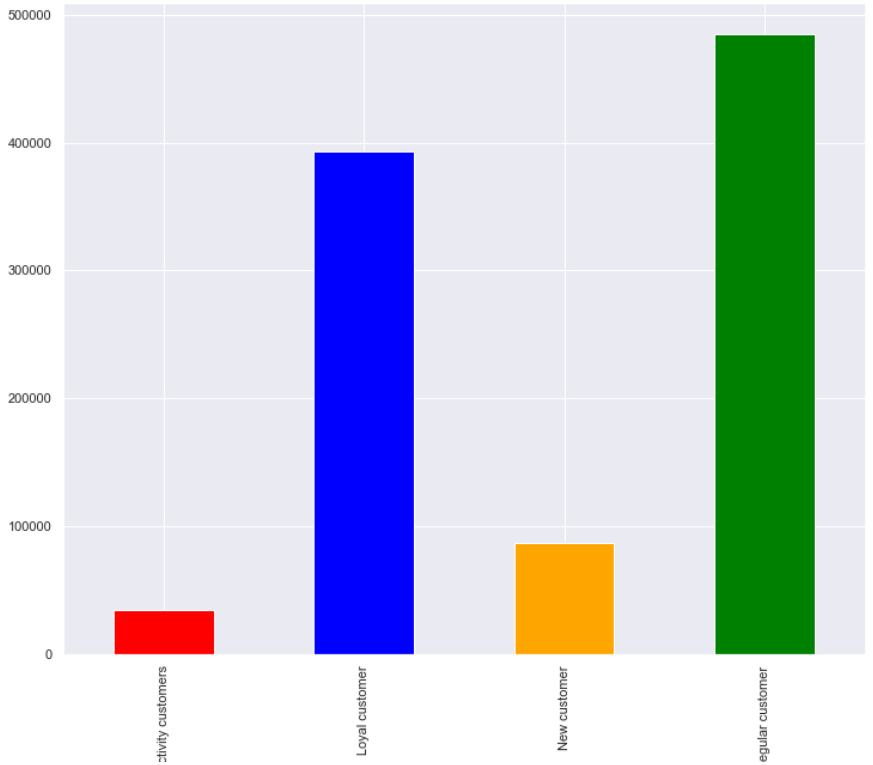
- **Objective:** You're an analyst for an existing company, Instacart, an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns.
- My task is to perform an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.
- The data set used is: "The Instacart Online Grocery Shopping Dataset 2017", Accessed from https://www.instacart.com/datasets/grocery-shopping-2017 on.

- Analysis has been conducted using Jupyter notebooks and the Anaconda libraries manager.
- I used Python and relevant libraries (pandas, NumPy, os, matplotlib, scipy, and seaborn).
- All required libraries have been successfully installed and imported into each script.
- Python scripts are clean and easy to follow with headings and contents lists.
- Data has been cleaned. Duplicate data, missing data, and mixed-type columns have been checked and addressed.
- Samples have been exported whenever an exclusions flag has been created.
- Any new columns that have been derived are relevant to the needs of the analysis.
- At least 4 types of data visualizations have been generated to communicate insights to stakeholders. Visualizations are clearly labeled.
- Data ethics have been kept in mind when dealing with data, especially in regards to customer information.

Relevant graphic representations:



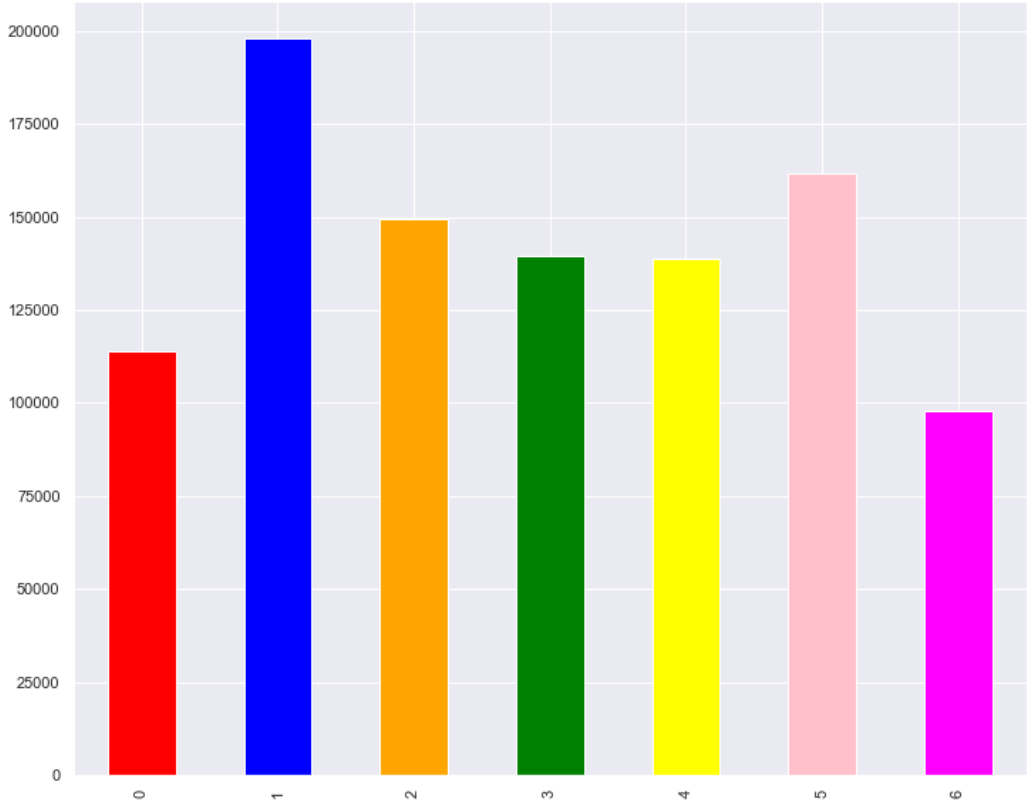
Customer profile



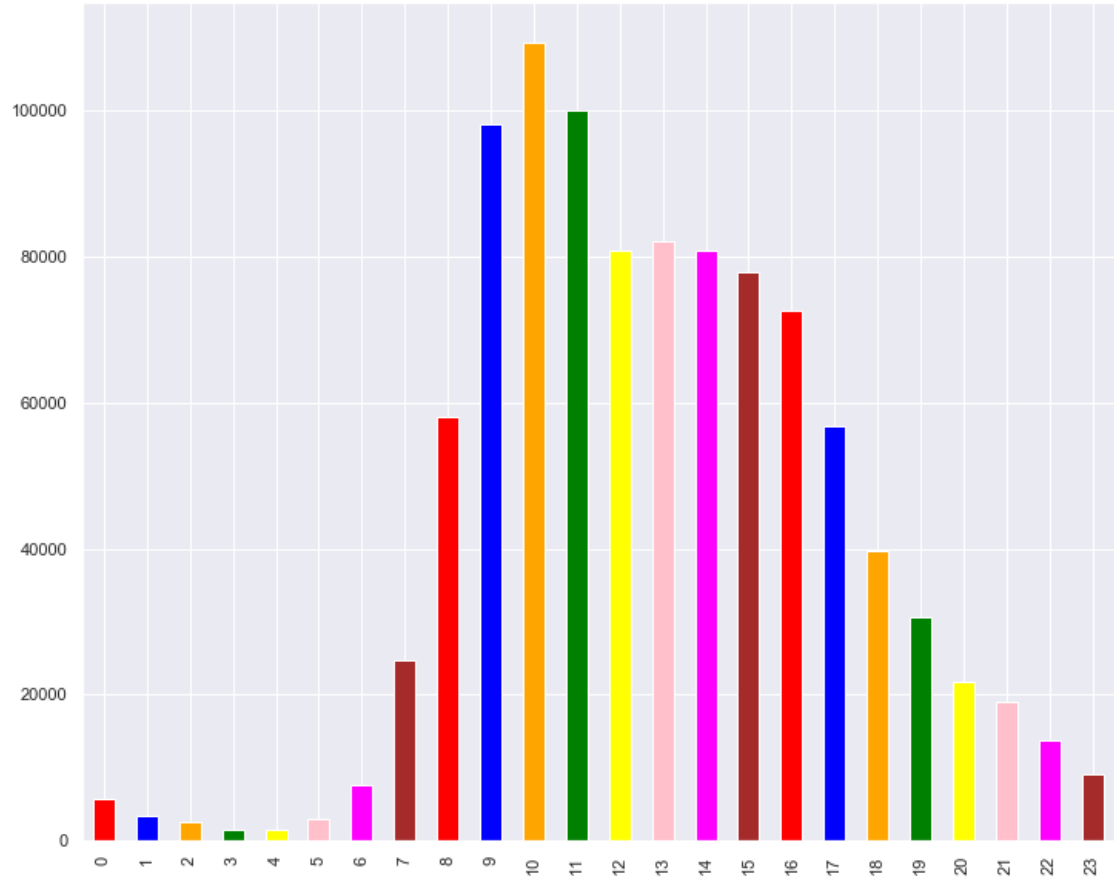
Loyalty flag

Source: My own realization in Jupyter

Relevant graphic representations:



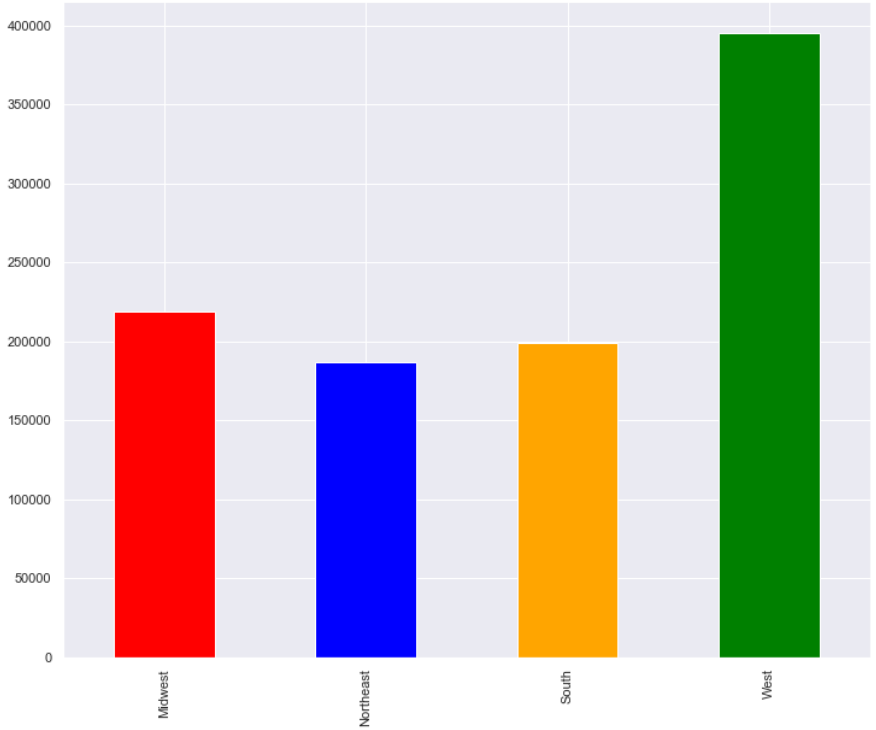
Orders days of week



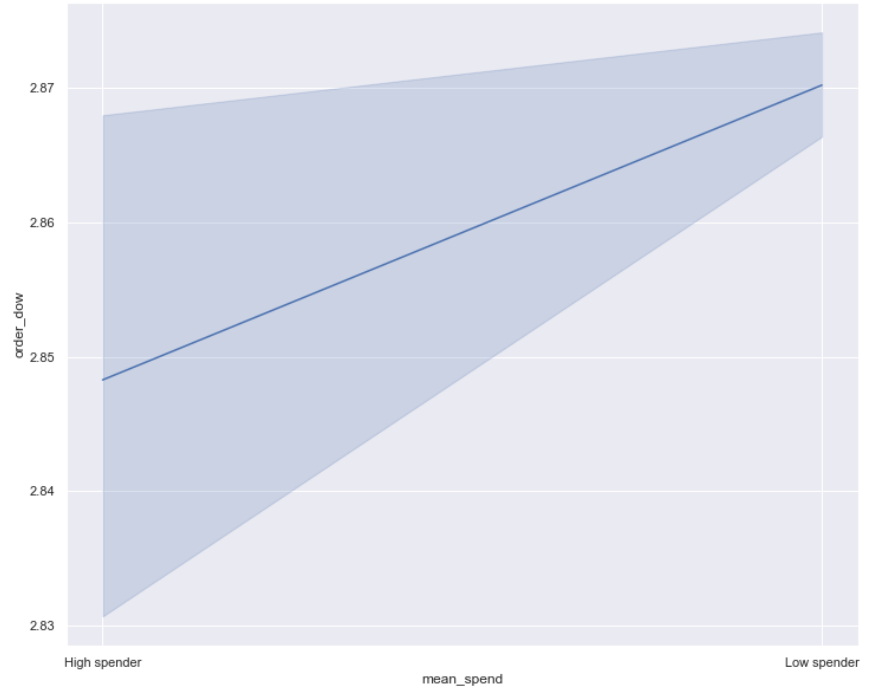
Orders hours of day

Source: My own realization in Jupyter

Relevant graphic representations:



Regions



Orders days of the week - mean spend

Source: My own realization in Jupyter

Conclusions and recommendations

- The sales team needs to know the busiest days of the week and hours of the day.
- Number of orders placed daily, we can see busiest days for Instacart is Saturday and Sunday.
- From 8 am to 17 pm the customers place orders. After 17 pm they are home with family. In the night the orders are only a few, this is normal.
- Find out from where the new customers find about the Instacart.
- There is a connection between the orders of days of the week and mean spendings of customers.
- **The Instacast should continuing to do marketing companies so that the customers to see different offers made by Instacart**.
- **Find out why the West region has so many customers in comparison with the rest. Is connected with the budget of marketing for each region.**

P5.Anti-money laundering projects at a global bank: Data Ethics and Applied Analytics

- **Objectives:**
- The characteristics of big data, how data analysts use big data, and the challenges of extracting knowledge from big data
- The impact of data bias and ethics on how data is used, shared, collected, and protected
- The fundamentals of data mining, including techniques for data mining and how it drives decision-making
- Predictive analysis and models such as linear regression
- Time-series analysis and time-series forecasting
- The basics of GitHub and how you can use it to refine your skills, collaborate with colleagues, and display a portfolio of work
- What to include in your portfolio when applying for jobs.

- The ways in which data is collected, used, and shared can be harmful to both individuals and society. Data collected on individuals, in particular, comes with responsibility. For these reasons, the data analyst should be guided by a strong ethical foundation and be able to discuss ethical concerns with their coworkers and employer.
- Besides being aware of data ethics and knowing how to raise ethical concerns with stakeholders, the data analyst needs to know how to derive useful information from big data. This is where data mining, predictive analytics, and time-series analysis and forecasting come into play.
- The data analyst needs to understand GitHub.

• BANK:

- **There is a cultural bias:** Mexican who work legally in the United States get paid in US Dollars but need Pesos to spend while they are home.
- In the background of the scenario its mentioned that Mexican drug cartels face the challenge of moving their ill-gotten US Dollars into Mexico. While the results of testing show that only 11 percent of customers whose transaction activity fits logic are Mexican Citizens. The 75 percent of the work items flagged suspicious involves Mexican Citizens.
- I would double-check from the data is coming.
- Who is the provider of the data? Are they trustworthy or not.

- The data that the team leader already has on ATM transaction from a project they led a few years ago could not be 100% biased free. So, I would suggest biased testing on ATM transaction data.

- If the investigators are **not trained properly**, they could have a huge impact on the accuracy of the testing results.
- And also if they don’t have **the minim of experience** this could impact negatively the accuracy of the testing results.

- **Whore is the persons that exited from the bank:** female, age between 34 – 51 years, from France or Germany, with balance 0, with salary between 0 – 50000 or 100000 – 150000, with credit score medium to high, not active members, without credit card and only with one product from the bank and with 9 or 8 tenure.
- **Why they left from the bank: they are from another country, they have a balance 0, they are not active members of the bank and they have only one product off the bank.**

	Credit Score	Age	Tenure	Balance	Number Of Products	Has Credit Card?	Is Active Member	Estimated Salary	Exited From Bank?
Average	636	45	5	90239.22	1	1	0	97155	Yes
Max	850	69	10	213146.2	4	1	1	199725	Yes
Min	376	2	0	0	1	0	0	417	Yes
Median	643	45	5	112434	1	1	0	100375	Yes
Average	651	37	5	74830.87	2	1	1	98943	No
Max	850	82	10	197041.8	3	1	1	199661	No
Min	411	2	0	0	1	0	0	371	No

Conclusions

- For bank:
- This job is to provide analytical support to its anti-money-laundering compliance department. This will involve a variety of data-related projects that help the bank assess client risk and transaction risk, as well as reporting on metrics.
- l helped build and optimize models that assist the bank in running their compliance program more efficiently.
- Many challenges is testing my technical skills and ability to handle data-related ethical dilemmas.

Thank you!