Jackson Lanier

Data Analyst

Project Case Studies

Projects

01 Evaluating historical sales data to understand the global market

02 Optimizing resource allocation based on demand forecasting

03 Identifying high CLV in preparation of a platform launch

04 Improving market segmentation by discovering growth opportunities

05 Providing analytical support to strengthen customer retention

06 Exploring an international market to understand factors impacting price

Evaluating historical sales data to understand the global market

Project Objectives

- Analyze historical data on video game sales to understand the global market
- Provide recommendations to GameCo, a fictional video game company, on the distribution of their marketing budget

Data Overview

- Historic video game sales data from 1980 to 2016
- Includes games that sold more than 10,000 copies
- Covers North America, the European Union, Japan, and other regions
- Sourced from VGChartz

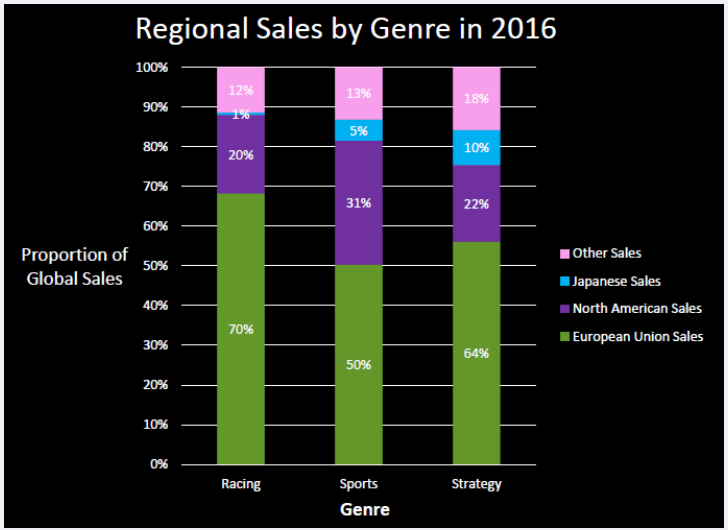
Techniques Applied

- Cleaning, grouping and summarizing data using Excel
- Transforming data using pivot tables
- Conducting a descriptive analysis using summary statistics and data distribution
- Displaying insights and relationships with visualizations

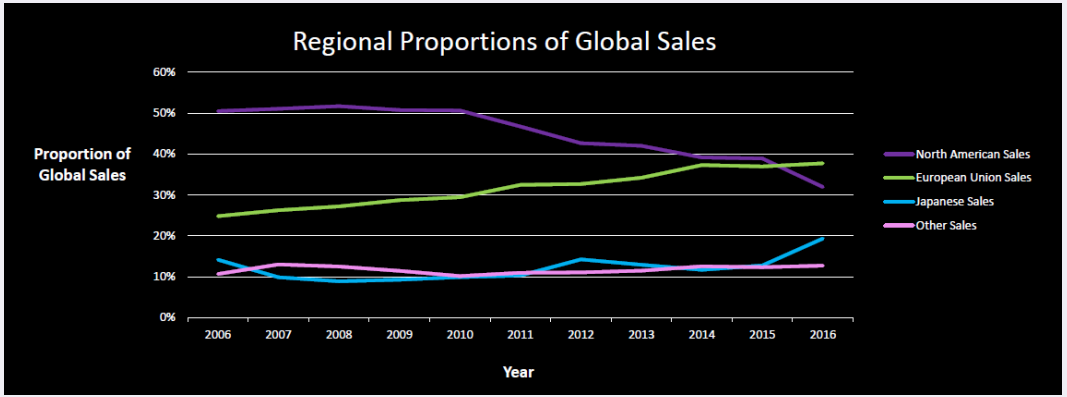
Tools Used

- Microsoft Excel
  - Statistics and distribution formulas
  - Pivot tables
  - Visualizations
- Microsoft PowerPoint
  - Presentation of findings

Visual Highlights



A stacked bar chart effectively displays how well racing, sports, and strategy games sold in the EU compared to other regions



The rising proportion of Japan's global sales should be taken into consideration when making global marketing decision

Key Recommendations for GameCo

- Focus their 2017 marketing efforts on how different genres perform in various regions
- Investigate cultural changes or phenomenon in each region to better understand why certain genres sell more units in specific regions

Actionable Insights

- Excel's ability to quickly clean, group, and summarize data should be utilized when appropriate
- Pivot tables are useful for transforming and aggregating data as well as creating visuals
- Quantifying a statement or discovery helps to provide meaning and understanding

Link to Final Presentation

Optimizing resource allocation based on demand forecasting

Project Objectives

- Examine trends in influenza and how they can be used to proactively plan for medical staffing needs across the country
- Distill business requirements and requests into questions that can be answered with an analysis
- Source and curate the data to address these questions
- Analyze the data, draw conclusions, and formulate recommendations
- Present findings in an easily consumable format

Data Overview

- US influenza deaths by geography from 2009 to 2017
- Sourced from the CDC
- US population data from 2009 to 2017
- Sourced from US Census Bureau

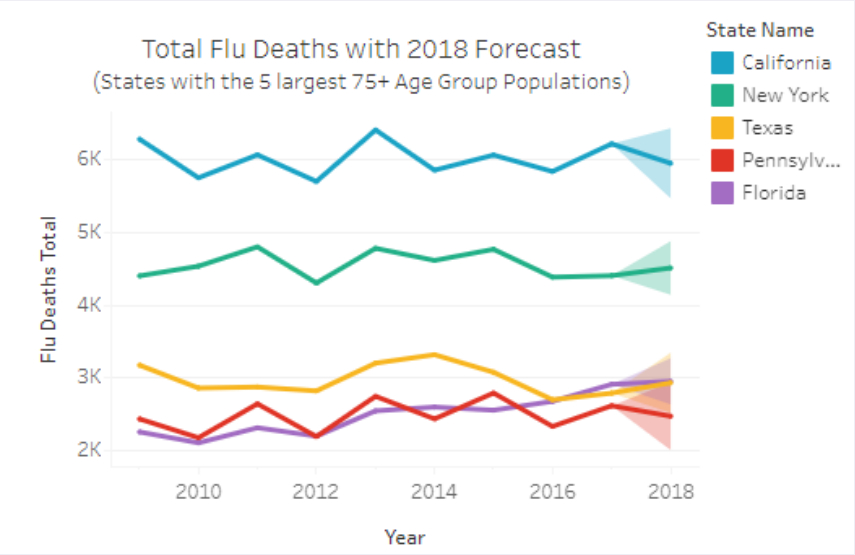
Techniques Applied

- Sourcing and describing datasets
- Profiling data, address data integrity issues, and implement data quality measures
- Calculating variance and standard deviation and testing for correlation between variables
- Performing statistical hypothesis testing
- Creating temporal, statistical, and spatial visualizations
- Forecasting a time series
- Narrating a cohesive presentation of findings and recommendations

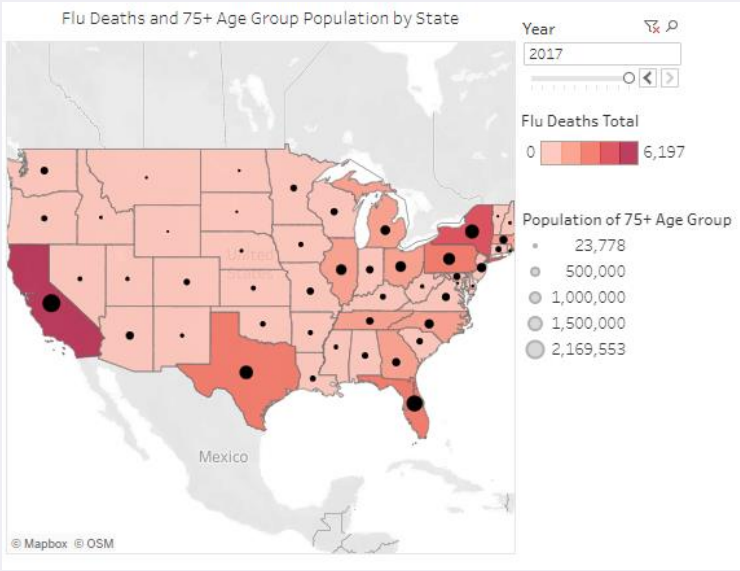
Tools Used

- Microsoft Excel
  - Data transformation and integration
  - Hypothesis testing
- Tableau
  - Visualizations
  - Forecasting
  - Dashboards
  - Storyboards

Visual Highlights



Medical staffing should prepare for thousands of flu victims in states with high populations of the 75+ age group



This map displays a connection between states with high counts of flu deaths and large 75 and older populations

Key Recommendations for Medical Staffing

- Data should be collected throughout the 2018 flu season so that an assessment can be made on the effectiveness of sending additional medical staffing to the states with the largest populations of the 75 and older age group.
- The medical staffing agency should consider providing info on the number of staff available in order to prioritize the CDC's recommendation of a 1-to-5 nurse to patient ratio.

Actionable Insights

- Data can be found from multiple sources, transformed, and integrated to create more comprehensive datasets.
- Correlation and statistical hypothesis testing are useful tools in measuring the strength of a relationship between variables

Link to Tableau Storyboard

Link to Video Presentation

Identifying high CLV in preparation of a platform launch

Project Objectives

- Assist Rockbuster Stealth, a fictional movie rental company, with the launch of their online video rental service
- Load all of Rockbuster's data into a relational database management system
- Use SQL to analyze the data
- Answer business questions from Rockbuster's business intelligence department
- Present findings in an easily digestible format

Data Overview

- Relational database
- Contains 15 tables on Rockbuster's film inventory, customers, payments, and more
- Downloaded from CareerFoundry

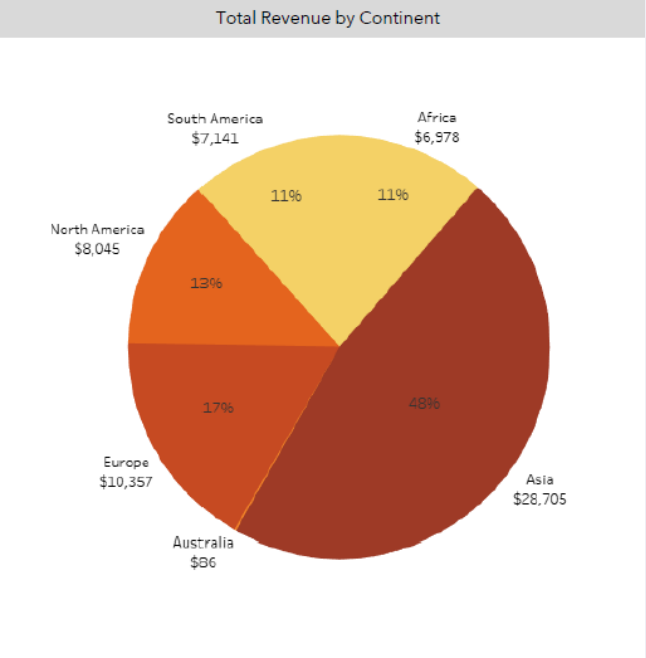
Techniques Applied

- Extracting an entity relationship diagram from a relational database
- Creating a data dictionary
- Writing SQL queries to sort, clean, filter, and profile data
- Joining tables with SQL
- Using subqueries and CTEs to answer moderate to complex business questions
- Visualizing and presenting SQL results

Tools Used

- PostgreSQL
  - Importing a database
  - pgAdmin 4 SQL query tool
- DbVisualizer
  - Visualizing Rockbuster's database
- Tableau
  - Visualizations
- Microsoft Excel
  - Compiling SQL query results
- Microsoft Word
  - Writing data dictionary
- Microsoft PowerPoint
  - Presentation of findings

Visual Highlights



48% of revenue is generated by Asia

Average Rental Duration	Total Number of Customers
• 4.99 Days	• 599
Average Rental Rate	Percentage of Active Customers
• $2.98	• 97%
Total Number of Films	Number of Countries with Customers
• 1,000	• 108
Most Common Film Rating	Total Recorded Revenue
• PG-13	• $61,312

Summary statistics provide a quick overview of Rockbuster's inventory and customer base

Key Recommendations for Rockbuster

- Rockbuster should further investigate why certain films perform better than others. The current data offers no clear indicator as to what causes a top grossing film.
- Rockbuster should consider including Australia in a global ad campaign and break into this populus market.
- Rockbuster's high lifetime value customers should be rewarded for the revenue they bring to the company.

Actionable Insights

- SQL can be used in creative ways to combine and aggregate data stored in a relational database
- The importance of a data dictionary and database schema when working with a relational database can't be understated
- Using the combined power of tools such as an RDBMS, Excel, and Tableau allows analysts to record and present findings in an easily consumable format

Link to Final Presentation

Link to GitHub Repository

Link to Tableau Visualizations

Improving market segmentation by discovering growth opportunities

Project Objectives

- Perform an initial data and exploratory analysis of Instacart product, department, orders, and customer data
- Derive insights and suggest strategies for better market segmentation
- Wrangle, combine, group, and aggregate data in order to answer business questions
- Visualize and present data in a professional format

Data Overview

- Open-source product, departmart, and order data from Instacart via Kaggle
- Customer data created by CareerFoundry
- Samples over 3 million grocery orders

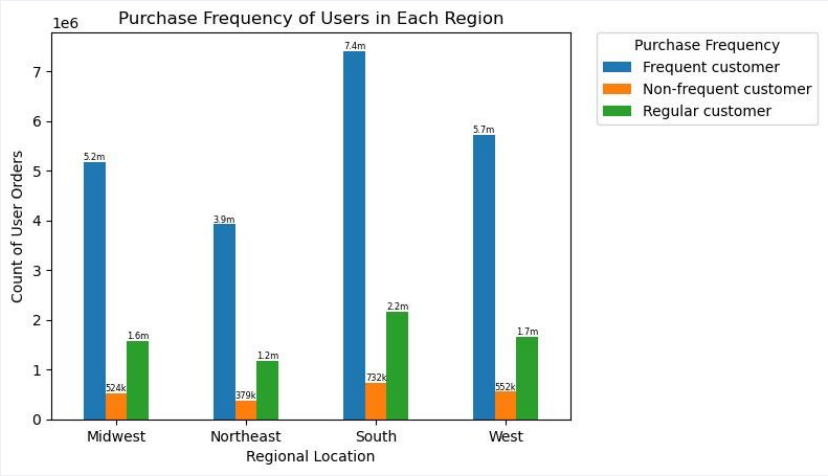
Techniques Applied

- Importing datasets into Jupyter notebooks
- Wrangling, cleaning, and combining datasets
- Deriving new variables for use in aggregation and grouping
- Exporting datasets in both CSV and PKL formats
- Visualizing variables and relationships between variables using Python to create histograms, bar charts, line charts, and scatterplots
- Presenting findings and recommendations

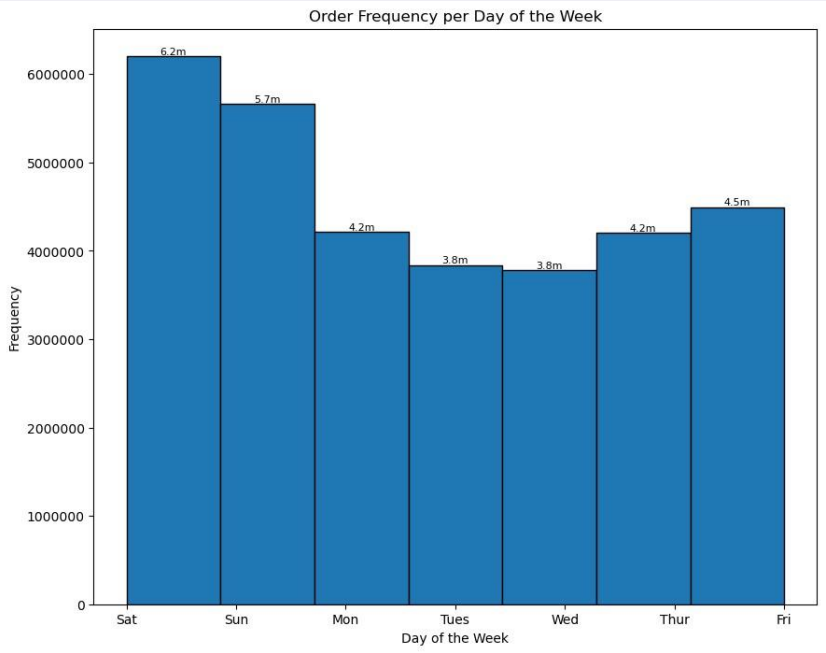
Tools Used

- Anaconda's Jupyter Notebook
  - Data analysis using Python
  - Python libraries including NumPy, Pandas, Seaborn, Matplotlib, and Scipy
- Microsoft Excel
  - Compiling, detailing and presenting analysis

Visual Highlights



Customer purchase frequency across regions is reflective of the customer count in each region



The middle of the week presents an opportunity for Instacart to increase order frequency

Key Recommendations for Instacart

- Instacart ads should focus on the convenience of Instacart deliveries during the week.
- Focus on advertising specific meals that use goods from a variety of departments. Similar to meal-kit deliveries such as HelloFresh or Home Chef, Instacart could advertise their ability to deliver a variety of specific goods that make up a meal.
- Instacart could potentially increase their loyal customer count by implementing a product recommendation system similar to other retailers such as Target and Amazon.

Actionable Insights

- Python is extremely useful for analyzing large datasets
- Importing Python libraries can greatly expand your abilities as an analyst
- Detailing scripts with notation allows others to understand your thought process
- Segmenting data into groups can help identify patterns and relationships that provide useful insights

Link to GitHub Repository

Link to Final Report

Providing analytical support to strengthen customer retention

Project Objectives

- Provide analytical support to the anti-money-laundering compliance department of Pig E. Bank, a fictional global bank
- Conduct basic data mining on a dataset and build a decision tree
- Create a time series and a simple moving average
- Identify and discuss ethical dilemmas and bias in relation to handling data in the workplace

Data Overview

- Pig E. Bank client data provided by CareerFoundry
- Includes info on client activity and demographics
- Historical data on oil prices provided by CareerFoundry

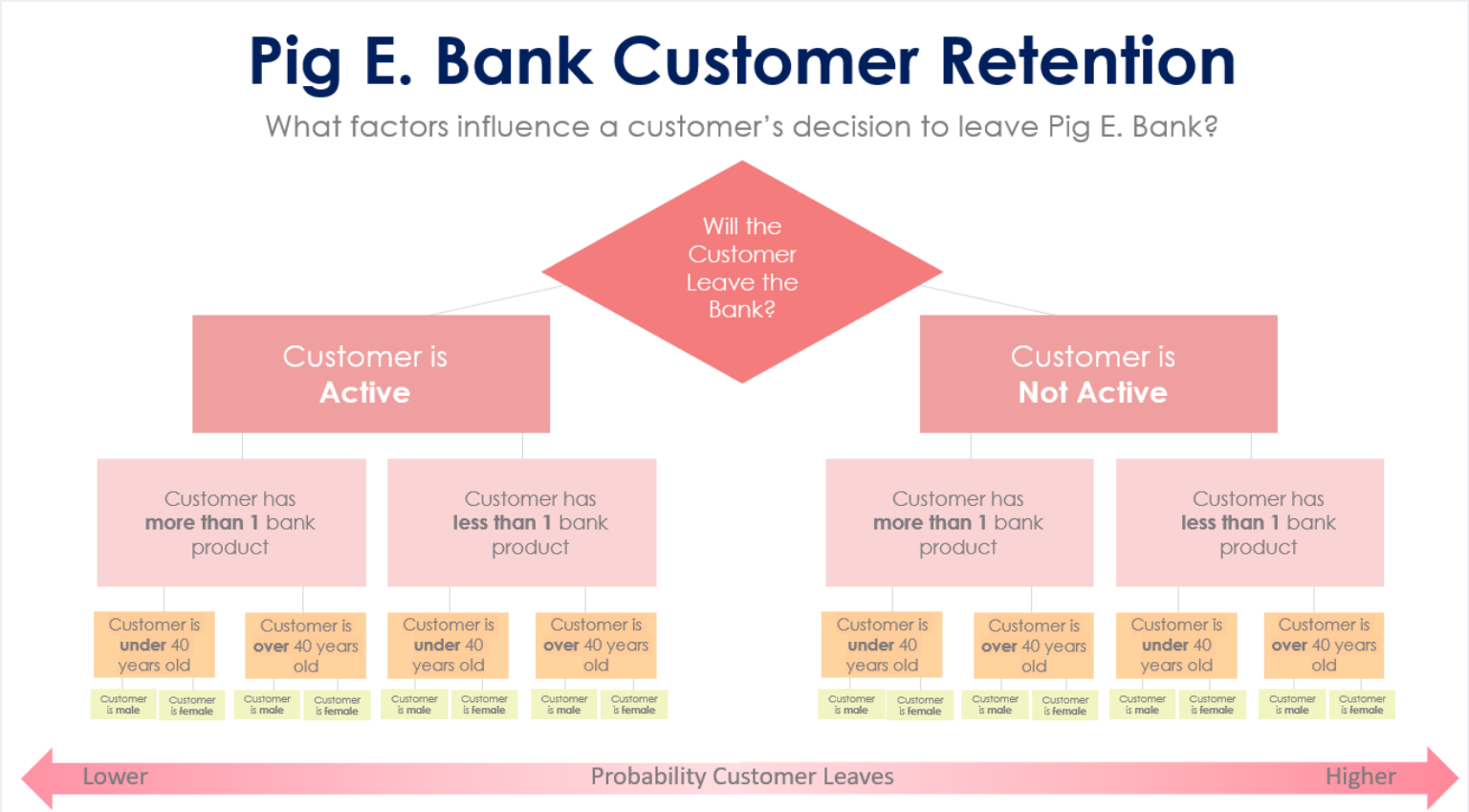
Techniques Applied

- Assessing data quality and cleaning data by addressing inconsistencies and missing values
- Documenting data quality and the data cleaning process
- Calculating descriptive statistics
- Determining factors that lead to client loss based on calculations
- Creating a simple moving average
- Identifying and discussing potential biases in data

Tools Used

- Microsoft Excel
  - Data cleaning
  - Documentation of the analysis process
  - Time series and simple moving average creation
- Microsoft PowerPoint
  - Decision tree creation

Visual Highlight



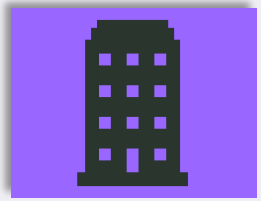
Non-active customers with less than 1 bank product who are over 40 years old and female are most likely to leave the bank

Key Recommendations for Pig E. Bank

- Customer retention efforts should start with customer activity. Perhaps Pig E. Bank can promote customer activity through incentives such as access to financial recommendations when deposits are made on a regular basis.
- Pig E. Bank could try advertising the benefits of their bank products and focus on the ease of use to target clients over 40.
- When handling sensitive customer data, Pig E. Bank should be careful to follow guidelines on data privacy and security. Not doing so risks exposing sensitive customer data.

Actionable Insights

- Converting raw data into a clean, quality data requires patience and an attention to detail
- Avoid making assumptions about what will be found in data. Let the data speak for itself
- Data bias and data ethics, including data privacy and security, are important factors to consider no matter what type of data you're working with.



Exploring an international market to understand factors impacting price

Project Objectives

- Source, clean, and analyze data from the web
- Use statistical analyses, including linear regression and cluster analysis, to explore relationships within the data
- Further explore the data through geospatial and time series analysis
- Create a Tableau storyboard to present relevant findings in a digestible format

Data Overview

- Portuguese real estate data sourced from Kaggle
- Historical residential property prices source from the Federal Bank of St. Louis
- Geographic data on Portugals region's sourced from simplemaps

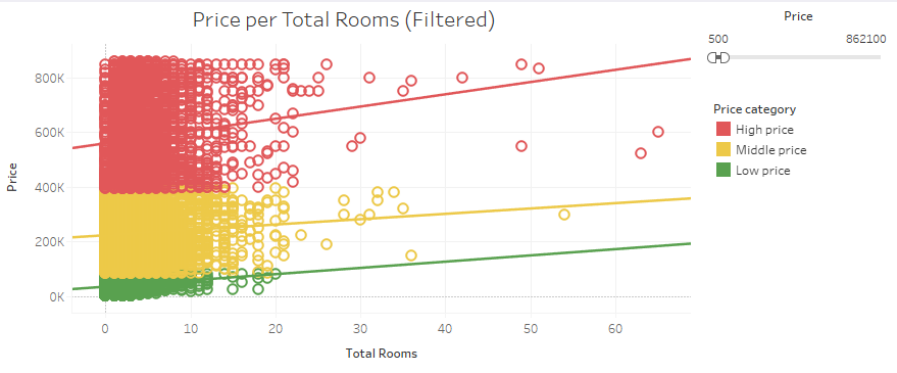
Techniques Applied

- Sourcing data suitable for analysis
- Cleaning data, including the use of machine learning to impute missing values
- Analyzing the relationships between variables within the data to identify the impact variables upon one another
- Forming and answering research questions related to the data
- Presenting relevant findings, including limitations and future possibilities

Tools Used

- Kaggle
  - Data Sourcing
- Anaconda's Jupyter Notebook
  - Data analysis using Python
  - Python libraries including NumPy, Pandas, Seaborn, Matplotlib, and scikit-learn
- Tableau
  - Visualization of analyses
  - Presentation of findings

Visual Highlights



The properties with the most total rooms tend to be in the high price range. However, there is no statistically relevant strength to the relationship between price and total rooms.

Lisboa, the coastal district containing Portugal's capitol, contains more available properties than any other district

Data Limitations and Future Analysis

- Limitations included the presence of null and irregular values in the original data as well as the massive variety of property types and range of variables present in Portuguese real estate.
- Exploring real estate comes with the general limitation that the market is constantly changing and is affected by a near endless number of factors
- Future analyses could dive deeper into the relationship between price and location or price and current economic conditions. Additionally, future analyses could explore the differences between asking prices and selling prices in the market.

Actionable Insights

- Data may not contain obvious relationships between variables. The lack of a strong relationship can be just as meaningful as a relationship with strength.
- The data cleaning process, such as removing outliers or imputing values, is a sensitive process. Comparing an analysis with and without the presence of outliers can be useful.
- Sourcing useable, relevant, available data can be a time-consuming process. The value of data can't be understated.

Link to GitHub Repository

Link to Tableau Storyboard

Please get in touch to discuss my portfolio and job opportunities

Email: jacksonlanier94@gmail.com LinkedIn: 

Don’t forget to take a look at my GitHub and Tableau Profiles