Danielle Copeland

Data Analysis Portfolio

About Me

I am a business analyst/project manager specializing in software development. Throughout my career, I have used tools such as Excel, Kibana, ElasticSearch, SQL and more to for data collection, cleaning, and analyzation. During my bachelor's and master's studies, I have used R and Stata to undertake statistical analysis.

My recent training with Career Foundry has increased my knowledge and skillset to make be a cutting edge data analyst.

Tech Skill Set

- Python (NumPy, Pandas, DataFrames)
- SQL
- R Studio
- Jupyter Notebooks
- Stata
- DbVisualizer
- Tableau
- Kibana
- Cognos
- Power BI

Portfolio Projects

LABOR FORCE PARTICIPATION
Analysis of the World Development Indicators to create recommendations of how to decrease the gender gap

INSTACART
Exploratory analysis on a customer database to derive insights on overall customer behavior

ROCKBUSTER
Business analysis on a movie rental company in order to launch a online video rental service

INFLUENZA SEASON
Analysis on influenza season to determine when and where to send staff to each state

GAMECO
Descriptive analysis on video game sales to inform the development of new games

P.E. BANK
Leverage data insights to advance the efficiency of an anti-money laundering model, using big data concepts

Labor Force Rate

Decreasing the world gender gap

Labor Force Participation Rate Background

OBJECTIVE

I wanted to learn more about the factors that influence the gender gap through an analysis of world development indicators.

KEY QUESTIONS

- Does poverty increase the percent of women?
- Does getting married or having children early decrease the percentage?

SKILLS & TOOLS

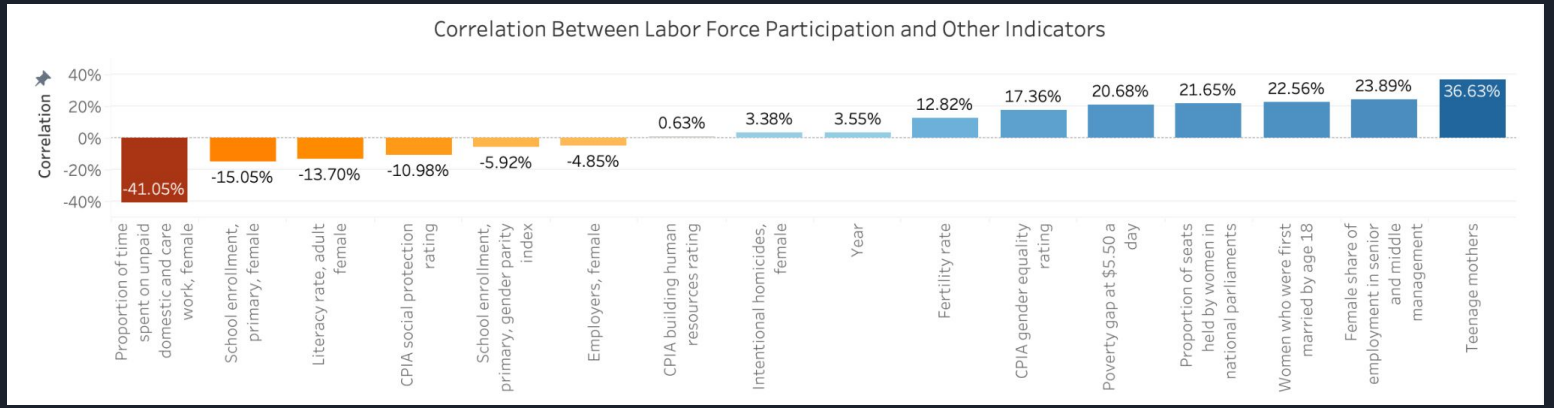
- Python (NumPy, Pandas, DataFrames)
- Data cleaning and wrangling
- Data grouping and aggregating
- Subsetting
- Tableau
- Anaconda
- Jupyter Notebook

Data Set - https://databank.worldbank.org/source/world-development-indicators/preview/on

Labor Force Participation Rate Analysis

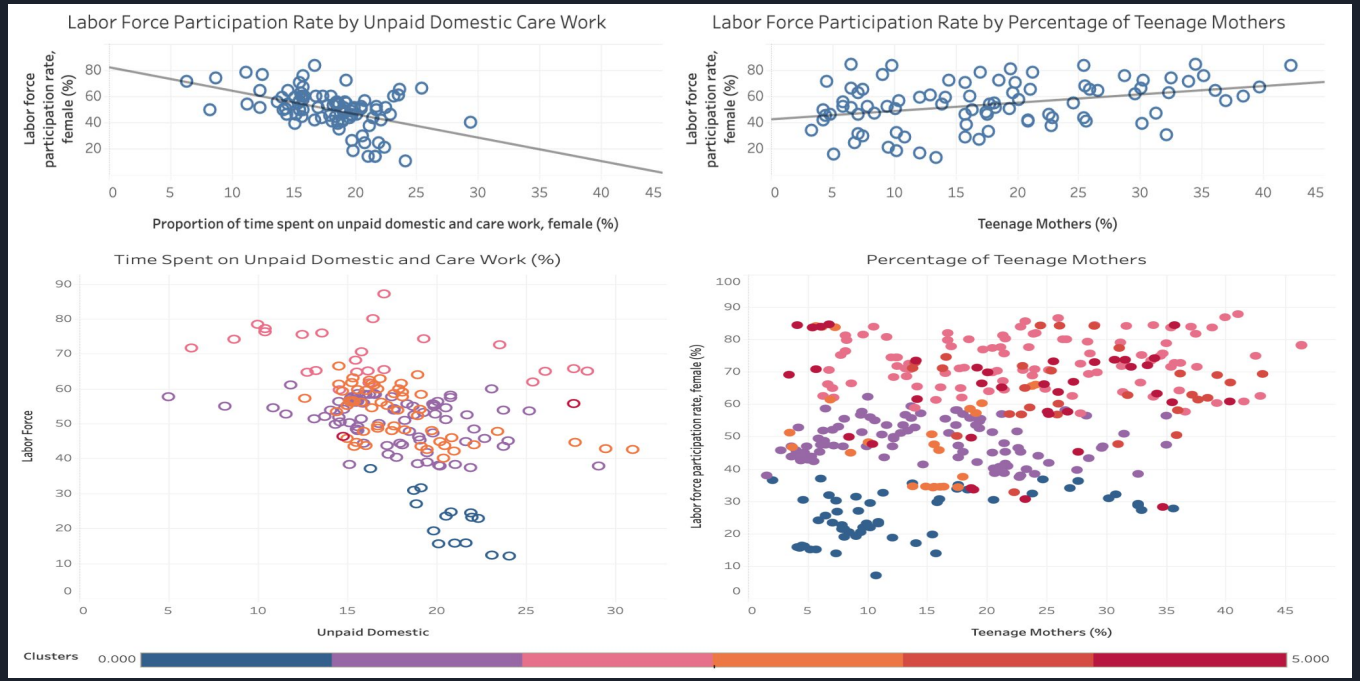
- There is a -41% correlation between the labor force rate and the proportion of time spent on unpaid domestic and care work
- The proportion of time can only explain about 20% of the labor force rate

- There is a 46% correlation between the labor force rate and the percentage of teenage mothers
- The percentage of teenage mothers can only explain about 13% of the labor force rate



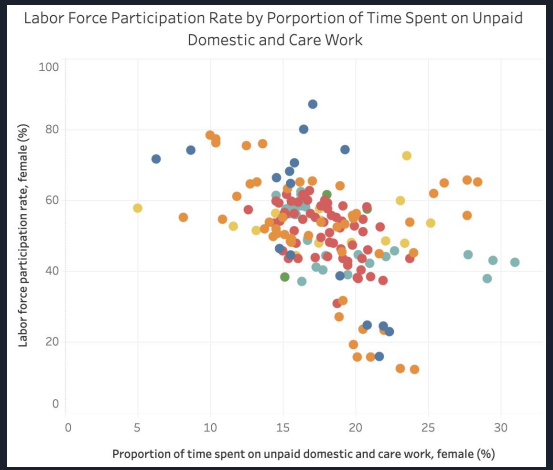
Linear & Cluster Analysis

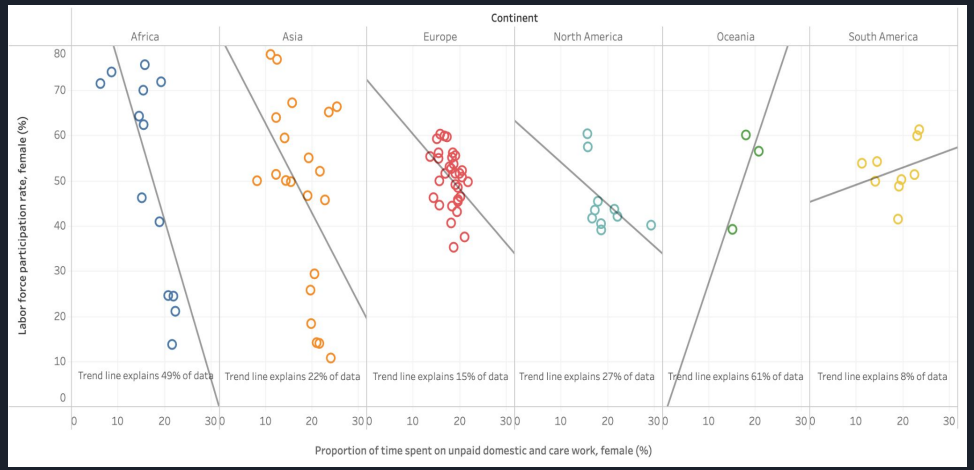
- The line for unpaid domestic care explains 20% of the data; the line for teenage mothers explains 13%.
- Through unsupervised machine learning, six subgroups were created.
- The number of clusters matches the number of continents, leading to the idea of grouping by continents.



Proportion of Time Spent on Unpaid Domestic and Care Work

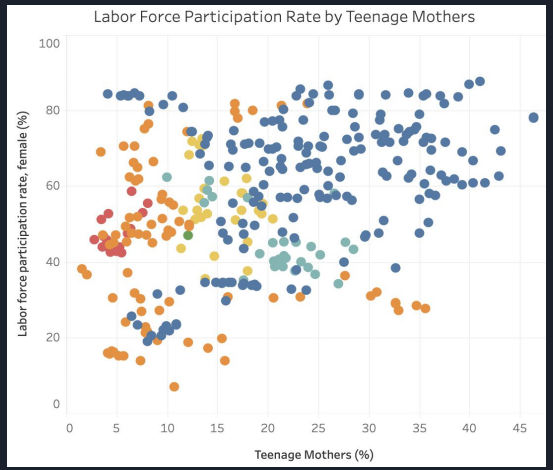
- Africa has the highest labor force percentage and the lowest unpaid domestic and care work percentage
- After breaking the data out by continent, the trends are more pronounced
- Surprisingly, Oceania and South America have a positive relationship

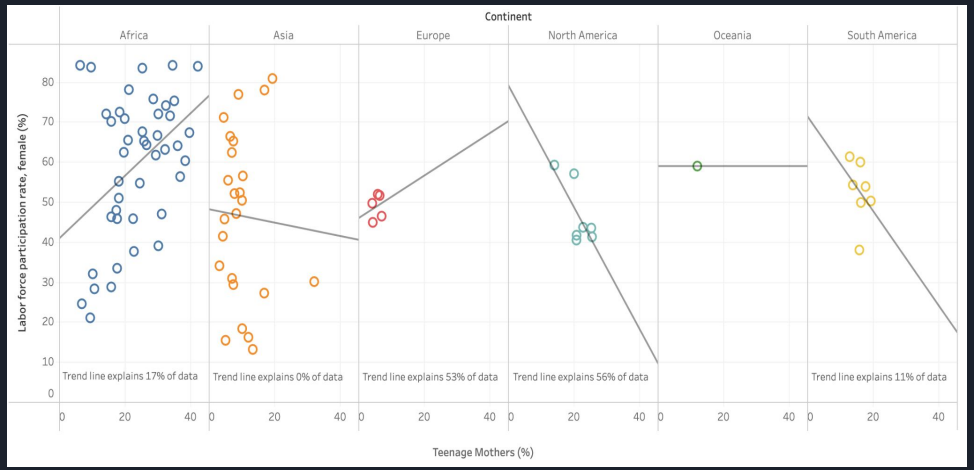




Percentage of Teenage Mothers

- Asia has no relationship between labor force and teenage mothers
- Africa and Europe has positive relationships while North America and South America have negative relationships
- There is not enough data for Oceania's relationship to be seen
- Overall, there can not be one general guidance for all continents as percentage of teenage mothers affects continents differently





Labor Force Participation Rate Recommendations

- Decreasing the proportion of time spent on unpaid domestic and care work may increase the LFPR in Asia, Africa, Europe and North America.
  - Fostering entrepreneurship could lead to women creating cleaning and care services that would decrease women's unpaid domestic and care work while allowing more jobs for a community.
- Decreasing the percentage of teenage mothers may increase the LPR in North America and South America.
  - Providing free birth control to young women may decrease the percentage of teenage mothers.

Tableau storyboard

Instacart

An online grocery store

Instacart Background

OBJECTIVE

An online grocery store wants to uncover more information about their sales patterns. They are interested in the variety of customers and their purchasing behaviors.

KEY QUESTIONS

- What are the busiest days/hours?
- What types of products are most popular?
- What types of customer profiles does the majority have?

SKILLS & TOOLS

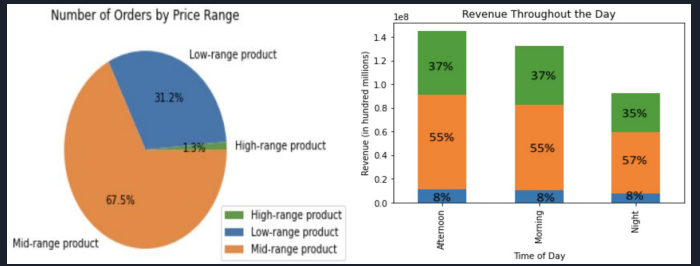
- Python (NumPy, Pandas, DataFrames)
- Data cleaning and wrangling
- Data grouping and aggregating
- Subsetting
- Jupyter Notebook
- Microsoft Excel

Data set - customer data set

Data set - Instacart online grocery shopping data set

Instacart Analysis

- The weekends are the busiest days of the week
- The middle of the day is the busiest
- Low and mid-range range products compose 87% of orders but only amount to 63% of revenue



Instacart Recommendations

- Grow
  - Create more brand loyal (shop more often) and customer loyal (make more purchases) customers
  - Grow the user base in the Northeast and Midwest as they spend the most per item.
  - Increase the number of high-range products
- Advertising
  - Run a high number of ads in the morning and night.
  - Advertise high-range products heavily
  - As 75% of customers have children, promote more kids/family friendly products.

Full Report

Rockbuster Stealth

A movie rental company

Rockbuster Stealth Background

OBJECTIVE

Rockbuster used to have stores around the world. They want to launch an online video rental service to stay competitive.

KEY QUESTIONS

- Which movies contributed the most/least to revenue?
- Which countries are customers based in?
- Do sales figures vary between geographic regions?

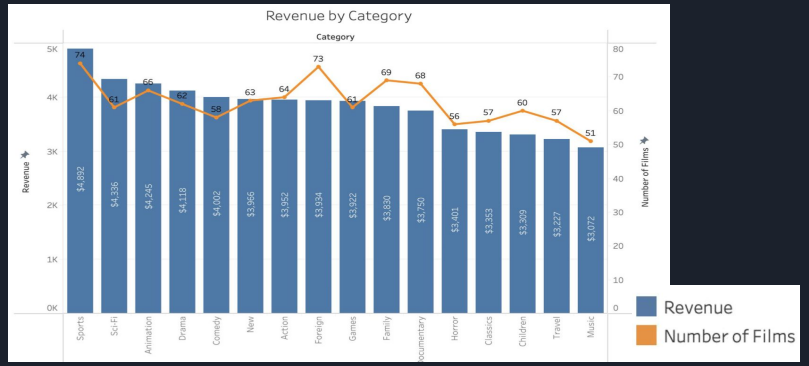
SKILLS & TOOLS

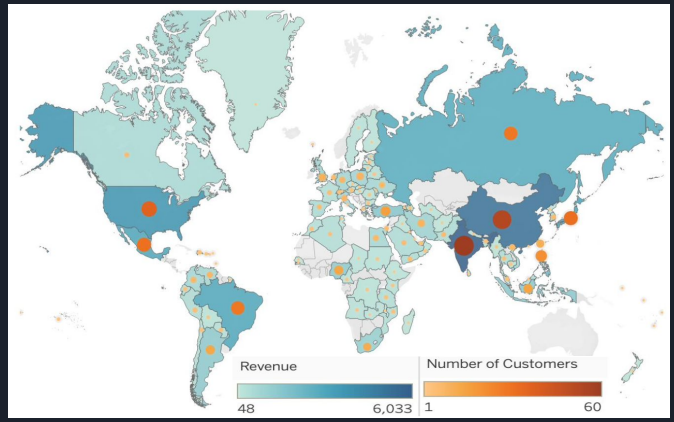
- DbVisualizer
- PostgreSQL
- Relational database management systems
- SQL queries
- SQL data cleaning
- Data dictionary

Data Set - Rockbuster data set

Rockbuster Stealth Analysis

- There is more inventory in categories generating less revenue
- Almost half of all customers (273) live in Asia
- Countries with more customers have higher revenue





Rockbuster Stealth Recommendations

- Diversify movies
  - Offload films that provide little to no revenue
  - Procure films from different release years
  - Increase the number of films in these categories: Sports, Sci-Fi, Animation, and Drama customer base
- Grow
  - Reward high lifetime value customers
  - Invest in films in different languages
  - Increase marketing in under represented geographical regions: North America, South America, Africa, and Oceania

Powerpoint presentation

Influenza Season

Medical staffing agency

Influenza Background

OBJECTIVE

The agency provides temporary workers to clinics and hospitals and wants to plan for staffing needs across the country. Demand for staff is highest during influenza season.

ASSUMPTIONS

Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital.

SKILLS & TOOLS

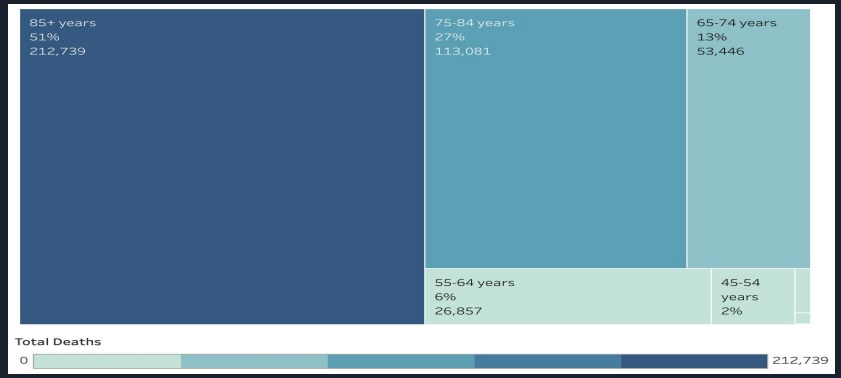
- Hypothesis testing
- Statistical analysis
- Data transformation
- Data integration
- Data sourcing
- Data profiling
- Tableau dashboards
- Visual Analysis
- Microsoft Excel

CDC - influenza deaths by demographics data set

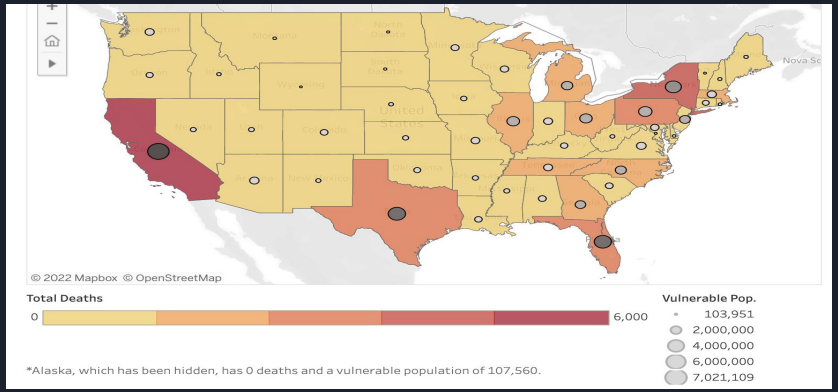
US Census Bureau - population by geography data set

Influenza Analysis

- Approximately 50% of deaths are people 85 years or older
- Over 90% of people are people 65 years or older



- States with larger populations have more deaths



Influenza Recommendations

- Additional staff members should be allocated based on the number of vulnerable people living in a state
- California, Texas, Florida, New York, and Pennsylvania have the largest number of vulnerable people (24.5-63.2 million) and deaths (2-6 thousand) so should have the highest support
- States with a large vulnerable population should ensure that these people are vaccinated against the flu

Tableau storyboard

GameCo

Video game analysis

GameCo Background

OBJECTIVE

Executives want to use data to inform the development of new games. Specifically, they want to better understand how their new games might fare in the market.

ASSUMPTIONS

Sales figures have remained the same between geographic regions over time. North America has the largest market share so investment marketing budget there.

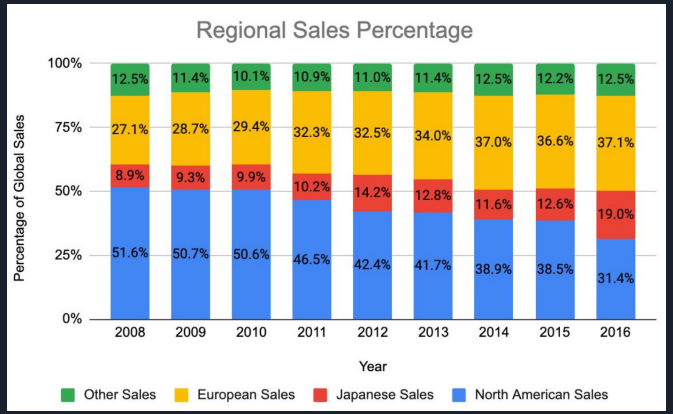
SKILLS & TOOLS

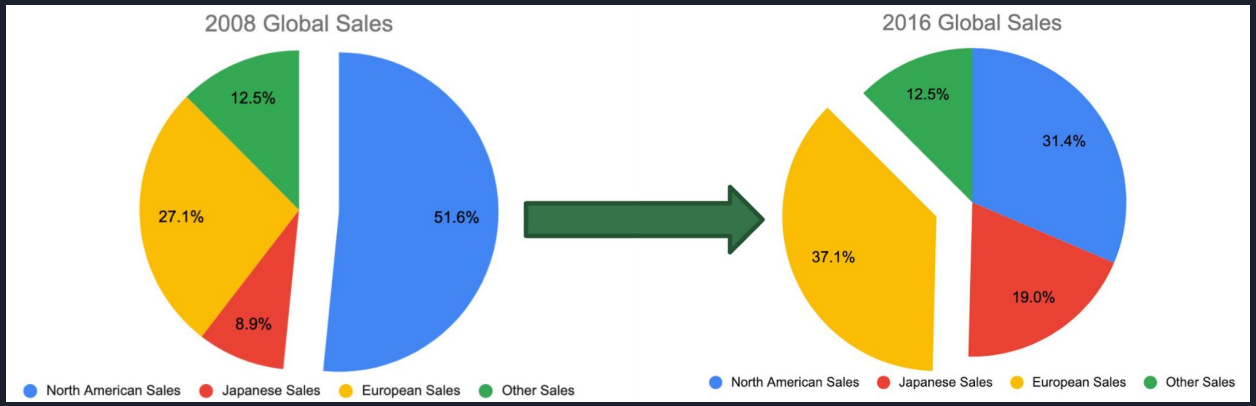
- Descriptive analysis
- Data visualizations
- Filtering/cleaning data
- Presenting results
- Pivot tables
- Microsoft Excel
- PowerPoint

Data Source: VGChartz
Data Set: video game sales data set

GameCo Analysis

- Since 2008, Japan and Europe have increased their market share while North American share has decreased
- The market leader has shifted from North America to Europe





GameCo Recommendations

- Invest in Europe and Japan
  - Regional sales have not remained consistent and the marketing strategy needs to reflect that
- Invest in Action, Shooter, Role-Playing, and Sports
  - The top 4 genres globally
- Create games in multiple platforms, especially PS3, X360, PS4, and 3DS
  - The top 4 platforms globally
- Keep the unique number of platforms to unique number of games ratio as high as possible

Powerpoint presentation

P.E. Bank

A global finance service company

P.E. Bank Background

OBJECTIVE

A well known global bank wants support for its anti-money-laundering compliance department. The bank needs to assess client and transaction risk.

KEY QUESTION

- What are the leading indicators that a customer will leave the bank?

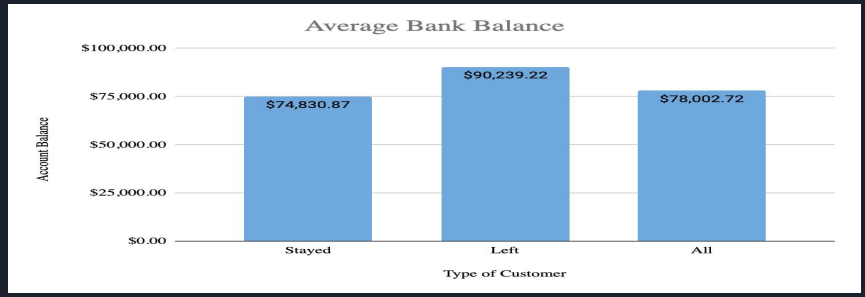
SKILLS & TOOLS

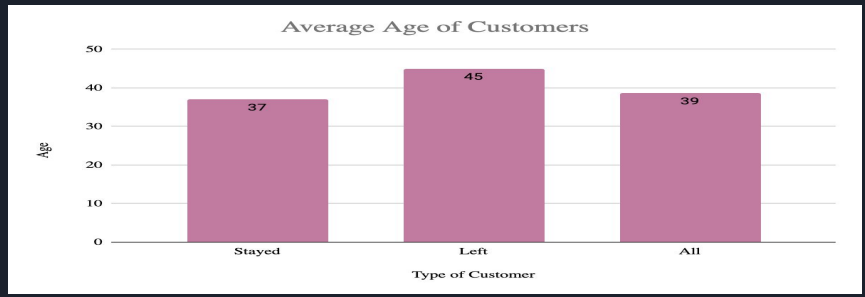
- Predictive analysis
- Time series analysis
- Data mining
- Forecasting
- Decision tree
- Data Security/Privacy
- Personally identifiable information (PII)

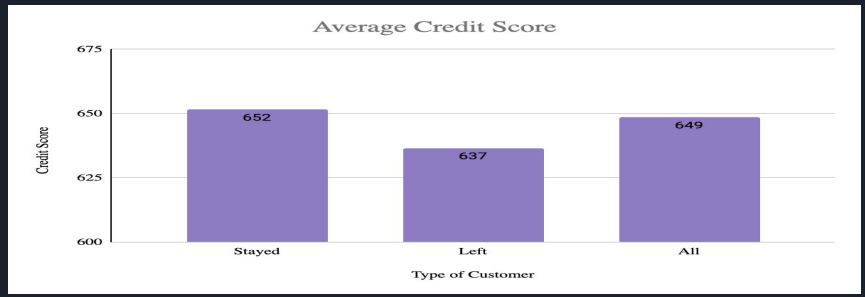
Data Set - client data set

P.E. Bank Analysis

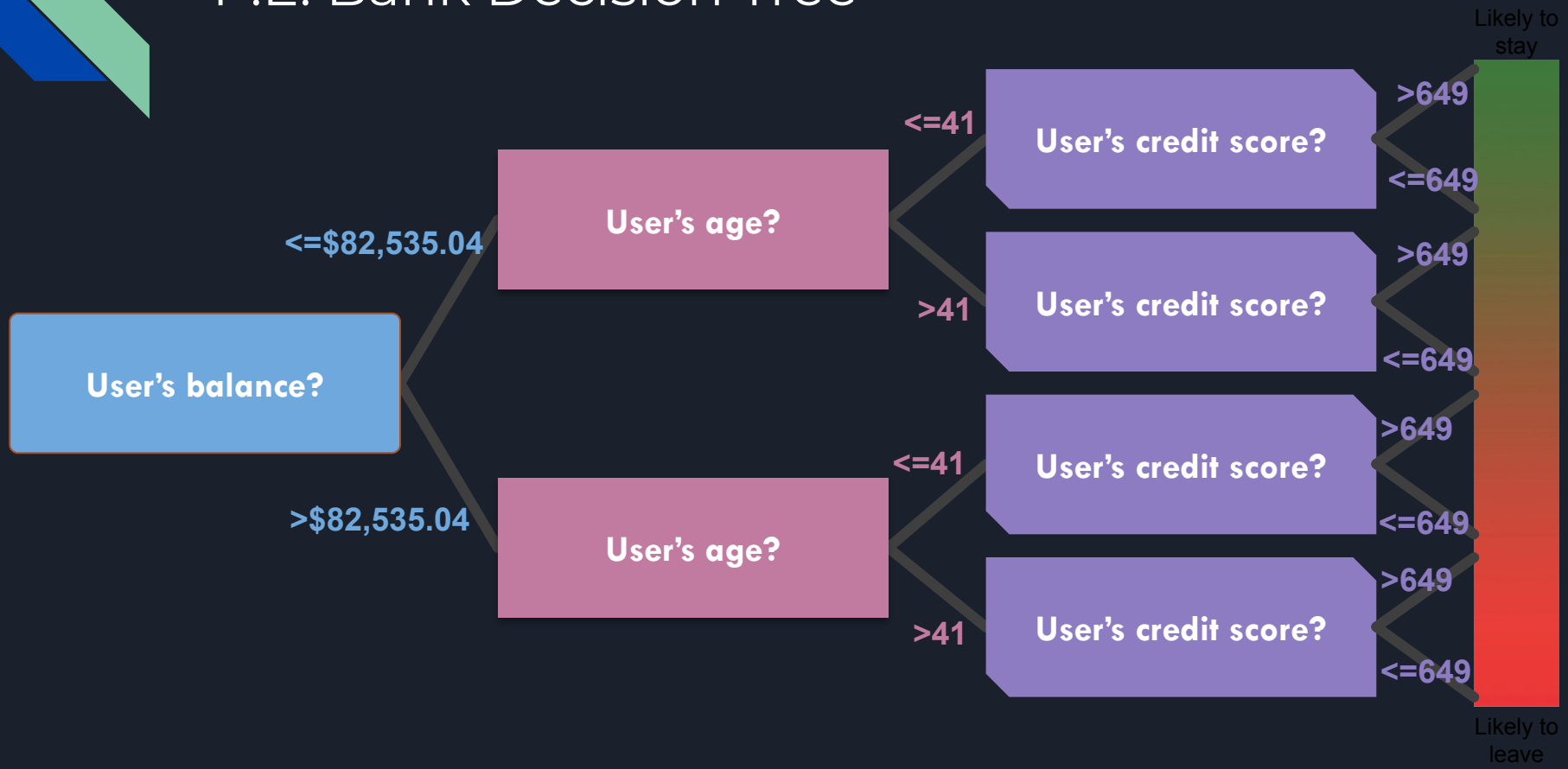
- Some variables have a large difference between the customers that leave, the customers that stay, and all customers.
- Three variables have statistically different means between those who leave and those who stay
  - Bank Balance
  - Age
  - Credit Score







P.E. Bank Decision Tree



Contact Me

Danielle Copeland

LinkedIn: https://www.linkedin.com/in/danielleocopeland/

GitHub: https://github.com/docopeland