

# Seda Aslan

Data Analyst Portfolio



# Projects

01

GameCo

02

Influenza season  
in the USA

03

Rockbuster  
Stealth

04

Instacart

05

PIG E. Bank

06

Global Air  
Pollution



01

# GameCo

Video Game Company  
Analysis

# GameCo

## Project Goal

- Understanding the historical trends
- Analyzing the current market situation
- Detecting the future market preferences
- Suggestion for budget allocation in 2017

## Objective

To perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market.

## Data Set

Historical sales of video games spanning different platforms, genres, and publishing studios. This data comes from the website [VGChartz](#).  
Video game sales data set.

## Tools



## Techniques

- Grouping & summarizing data
- Data sorting, filtering and cleaning
- Deriving new variables
- Descriptive analysis
- Visualizing result in Excel
- Presenting results

# Data analysis stages

## Data cleaning

- ✓ Remove duplicates
- ✓ Remove irrelevant data values
- ✓ Replace missing values
- ✓ Correct formatting inconsistencies



## Data grouping and summarizing

- ✓ Using pivot tables to transform more usable and insightful.
- ✓ Grouping data
- ✓ Filtering your data to look at specific subsegments
- ✓ Creating new variables to glean many insights



## Descriptive analysis

- ✓ Analyzing basic statistics features for central tendency; mean, median, mode
- ✓ Analyzing symmetrical and skewed distribution
- ✓ Analyzing the spread of data; ranges(quantiles)
- ✓ Identifying outliers

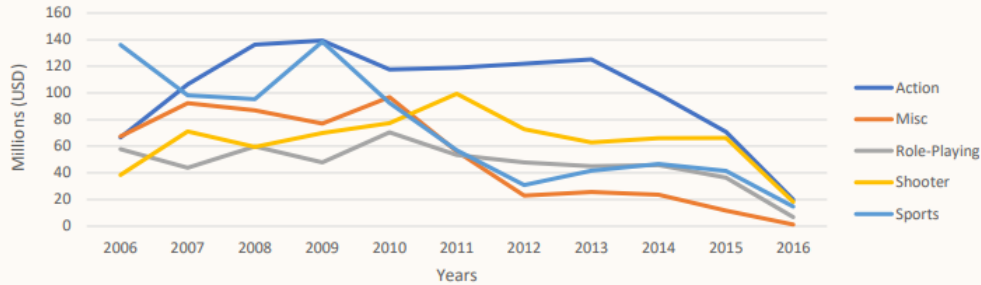


## Developing insights & visualizations

- ✓ Visualization based on data insights
- ✓ Analyzing of market trends according to key objectives
- ✓ Collecting of project deliverables (reflections, presentation)

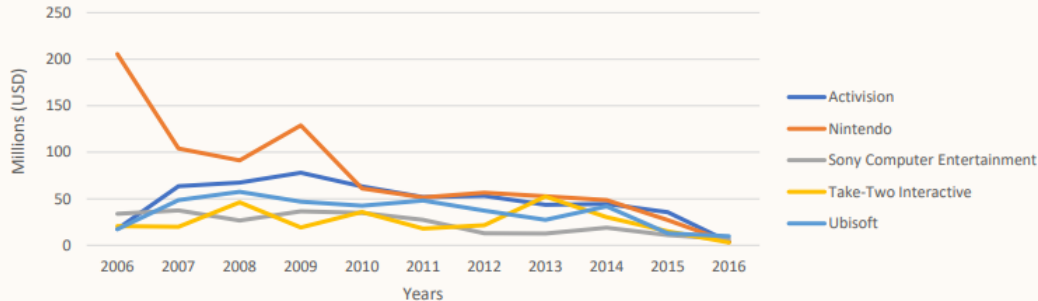
# Analysis of Genres and Publishers

Sales by genres last 10 years



There were decreases in the sales but Action games were always the most popular.

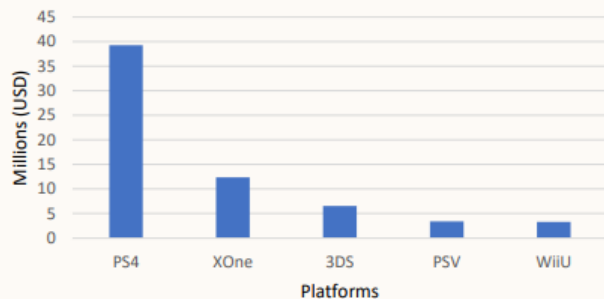
Sales by publishers last 10 years



There were decreases in the sales but Nintendo led the market last 10 years.

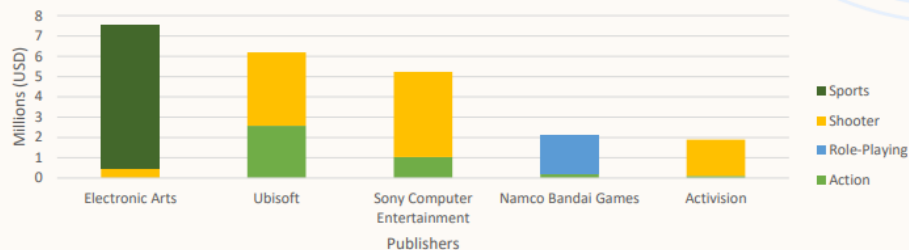
# Analysis of Platforms

Top 5 platforms – Global Sales

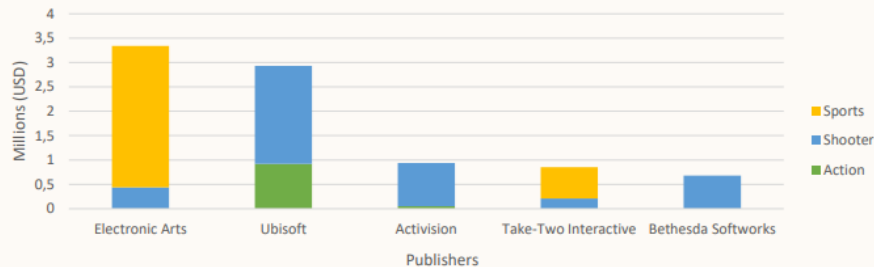


- PS4 is the most popular game platform.
- EA and Ubisoft have more PS4 games than other publishers.
- XOne also getting popular in the markets.
- Sports and Action games are more popular on PS4 and XOne platforms.

PS4



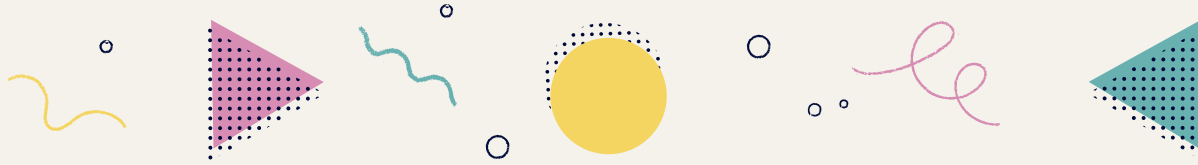
XOne



# Recommendations

- The recommended platforms are PS4, XOne, and 3DS to allocate more budget in 2017. These platforms have the big market shares and have increasing sales trends. That fact shows us that these platforms are the most popular in the market.
- Action, Shooter, and Sport games must have more budget to promote in NA and EU in 2017. These genres have more interest and increasing popularity in the markets.
- Japan market is different than other markets and need special attention. As the big competitors doesn't have footprints there, there is a great opportunity to grow with role-playing games.





## Additional Sources

Detailed Data Analysis



Final Presentation



Project Reflections





02

# Influenza season in the USA

Preparing for Influenza  
Season

# Influenza season in the USA

## Project Goal

To help a medical staffing agency that provides temporary workers to clinics and hospitals as needed. The analysis will help plan for the influenza season when additional staff is in high demand.

## Objective

Determining when to send staff, and how many, to each state. Creating forecast influenza season for 2018.  
Determining characteristics of vulnerable population and their locations.

## Data Set

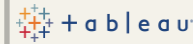
Influenza deaths by geography, time, age, and gender

Source: [CDC Data Set](#)

Population data by geography

Source: [US Census Bureau Data Set](#)

## Tools



## Techniques

- Data quality check, profiling and cleaning
- Data integration and transformation
- Data grouping and summarizing
- Statistical hypothesis testing (t-test)
- Forecasting
- Visual analysis
- Visualizing result
- Storytelling in Tableau

# Data analysis stages

## Planning a research project

- ✓ Formulate research questions and hypotheses
- Source relevant data for the project.
- ✓ Prepare a project management plan.

## Data Preparation

- ✓ Check for integrity and quality issues in the data.
- ✓ Clean data by removing duplicates and handling missing values.
- ✓ Transform and integrate data sets if necessary.

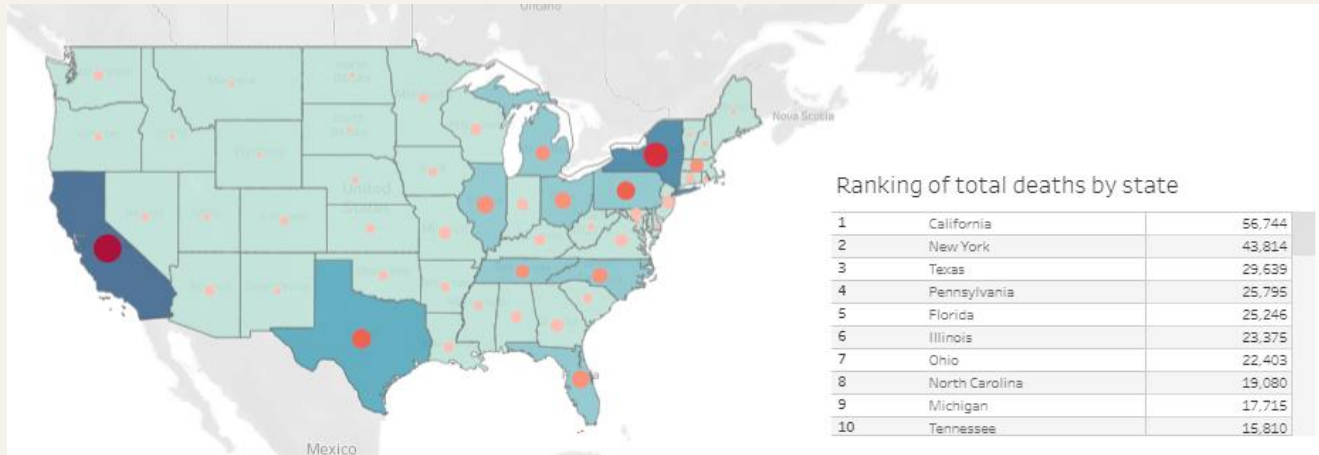
## Statistical hypothesis testing

- ✓ Calculate the variance and data spread for two variables.
- ✓ Calculate the correlation between the variables.
- ✓ Conduct a two-sample t-test.

## Data visualization & Storytelling in Tableau

- ✓ Create composition and comparison charts.
- ✓ Create temporal visualizations and forecasting.
- ✓ Create a narrative and record a video presentation.

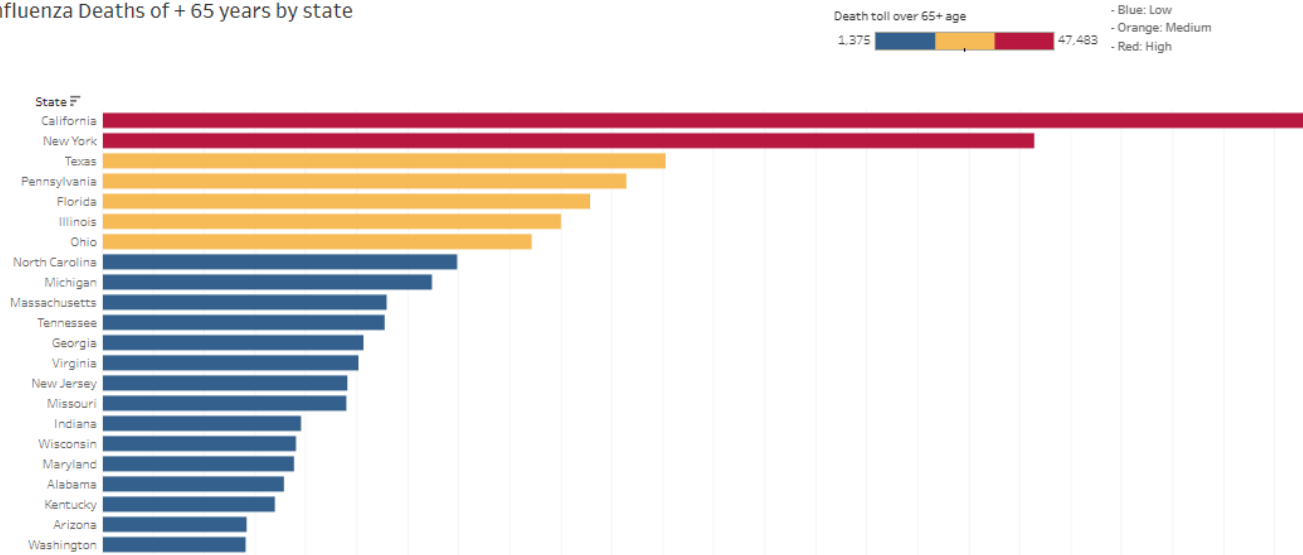
# Analysis of states by total deaths



These 10 states had very big death rates during the influence season but California and New York had the way more deaths than others. We clearly see here the numbers of deaths.

# Analysis of states by vulnerable group

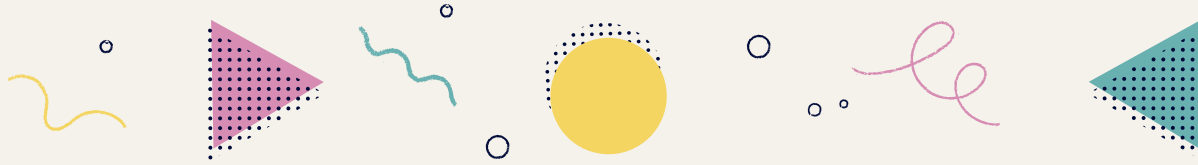
Influenza Deaths of + 65 years by state



These 10 states have big risk groups but the most emerging states are California and New York. There were big numbers of deaths during the previous seasons and most of these people were older than 65 years old. These states have very high vulnerability groups and it is highly possible to see very big need to medical staff.

# Recommendations

- High death rate months are December, January, February and March.
- The preparation should start at the beginning of September and health staff should be provided in mid of September.
- High death rate states are California, New York, Florida, Illinois, Ohio, Pennsylvania and Texas.
- There should be more staff provided as early as possible to high vulnerable states to mitigate the risk faster.



## Additional Sources

Detailed Data  
Analysis



Tableau  
Storyboard



Interim report and Project  
management plan



Video  
Presentation







03

# Rockbuster Stealth

International Business  
Analysis of Online Video  
Rental Services

# Rockbuster Stealth

## Project Goal

The goal of this exercise is to help Rockbuster Stealth, a movie rental company, launch an online video rental service by analyzing their existing data and providing actionable insights for decision-making.

## Objective

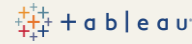
Identify top and bottom revenue-generating movies.  
Determine the average rental duration for videos.  
Analyze customer distribution by country.  
Identify customers with high lifetime value.  
Analyze sales variations between geographic regions.

## Data Set

### Rockbuster Data Set

*Data set contains data only for 2006*

## Tools



PostgreSQL



DbVisualizer

## Techniques

- Understanding relational databases
- Analysing Entity Relationship Diagram (ERD)
- Database querying and Filtering, cleaning in SQL
- Joining tables
- Coding common table expressions (CTE)
- Creating a data dictionary
- Visualizations of results in Tableau

# Data Analysis stages

## Database Analysis

- ✓ Set up a PostgreSQL database environment.
- ✓ Analyze keys and indexes in the database.
- ✓ Extract an Entity-Relationship Diagram (ERD).
- ✓ Begin creating a data dictionary.



## Cleaning & Summarizing data in SQL

- ✓ Perform basic CRUD operations and SQL commands.
- ✓ Organize, sort, and filter data.
- ✓ Identify and clean dirty data.
- ✓ Generate summary statistics for exploratory data analysis (EDA).



## Answering Business Queries

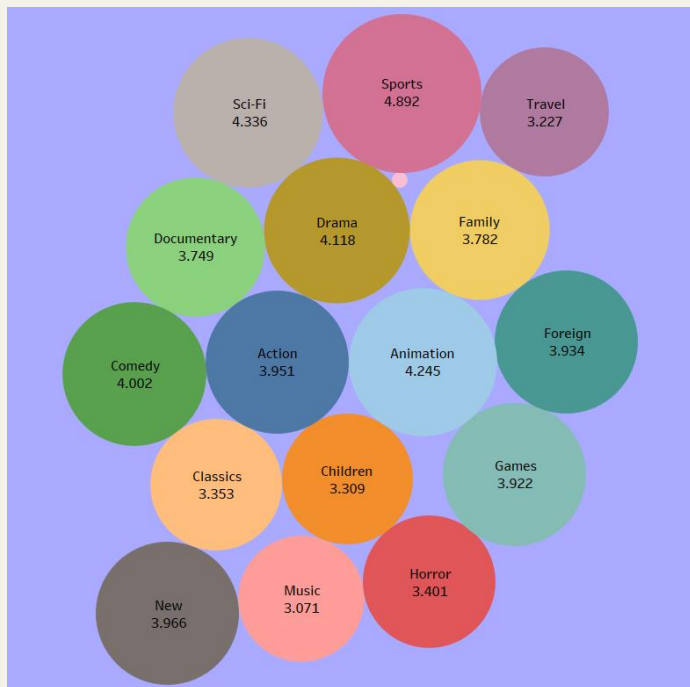
- ✓ Use SQL joins to combine tables.
- ✓ Write subqueries and apply common table expressions (CTEs).
- ✓ Answer complex business questions using CTEs, subqueries, and joins.



## Presenting results

- ✓ Create visualizations and a storyboard in Tableau.
- ✓ Prepare a final presentation with analysis results and recommendations in a PPT format.

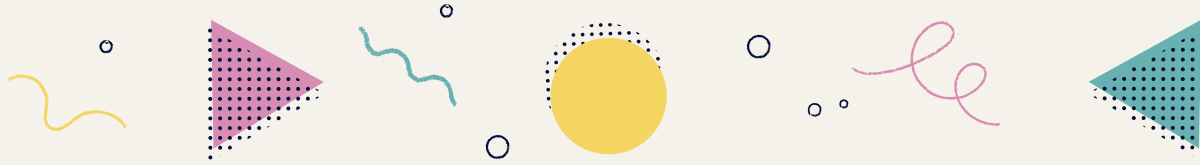
# Analysis of top categories



Sports is the most successful genre that Rockbuster offers. Sports produced the most income, totaling \$4,892k. However, Thriller was the genre that produced the least income, with only \$47k earned.

# Recommendations

- Instead of launching to all countries, Rockbuster should select the best performing countries based on analyses to start.
- Instead of starting wide, it is better to select the best performing genres, and ratings to accelerate faster.
- The extension to other countries, genres, and ratings should be considered based on the results of first launches
- As the user experience already exists, Rockbuster should have a proper subscription model to reach all the customer segments.



## Additional Sources

SQL code



Final Presentation



Data dictionary





04

# Instacart

Customer profile, buyer behavior,  
and sales analysis for an online  
grocery delivery service

# Instacart

## Project Goal

To segment customers, map popular products to each segment, provide personalized recommendations, test marketing campaigns, and maximize customer lifetime value. The goal is to improve customer engagement, increase sales, and optimize marketing efforts for Instacart.

## Objective

Determining busiest days and hours for sales. Using simpler price range groupings. Knowing which departments have the highest frequency of product orders. Determining customer ordering profiles based on various features such as brand loyalty, family status regions, etc.

## Data Set

- Customer
- Products
- Orders
- Departments
- Data dictionary

## Tools



## Techniques

- Data wrangling and subsetting
- Data merging
- Data consistency checks
  - Deriving variables
- Grouping data
- Combining and exporting data
- Visualizations in Python
- Excel reporting



# Data analysis stages

## Data Preparation

- ✓ Conduct basic descriptive exploratory
- ✓ Create new data frames based on a certain criteria
- ✓ Fix mixed-type variables
- ✓ Expose and deal with missing values
  - ✓ Expose and remove duplicates
- ✓ Merge the data frames

## Deriving new variables

- ✓ Create new columns using conditional logic
- ✓ Create flags and place them in new columns
- ✓ Create summary columns of descriptive statistics

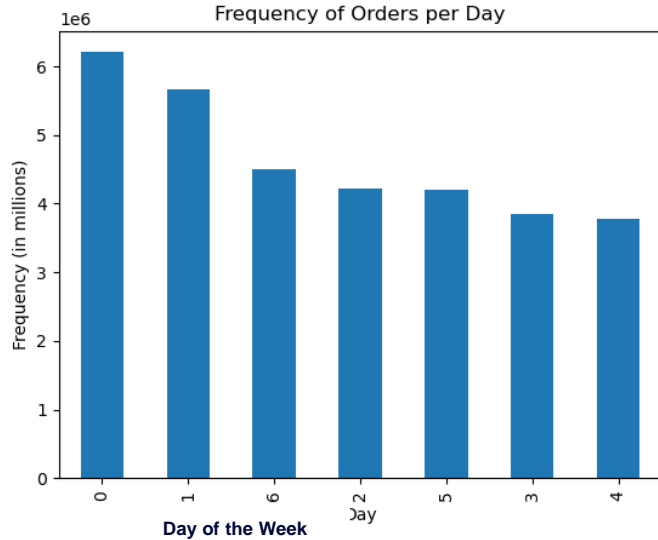
## Data visualization

- ✓ Create histograms, bar charts, line charts, and scatterplots for different variables and relationships between variables

## Excel reporting

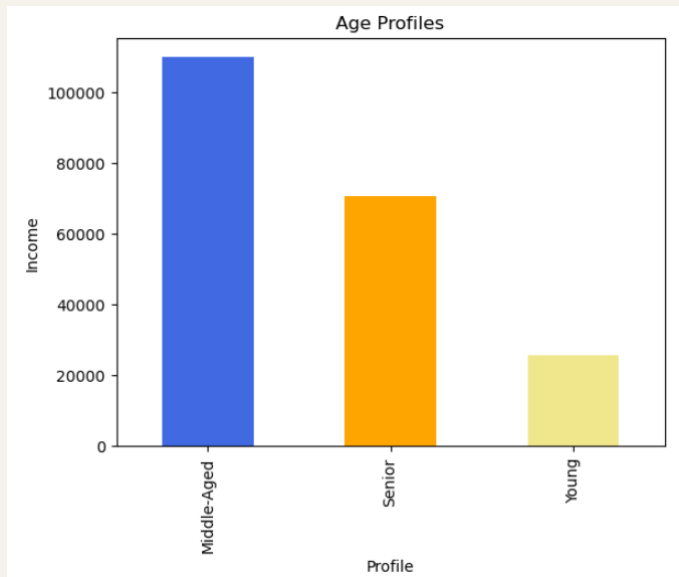
- ✓ Analyze order behavior of different customer groups
- ✓ Summarize analysis findings and describe what connections in the data you've found
- ✓ Create a report that describes the analysis methodology, the results, and the recommendations

# Analysis of Busiest days

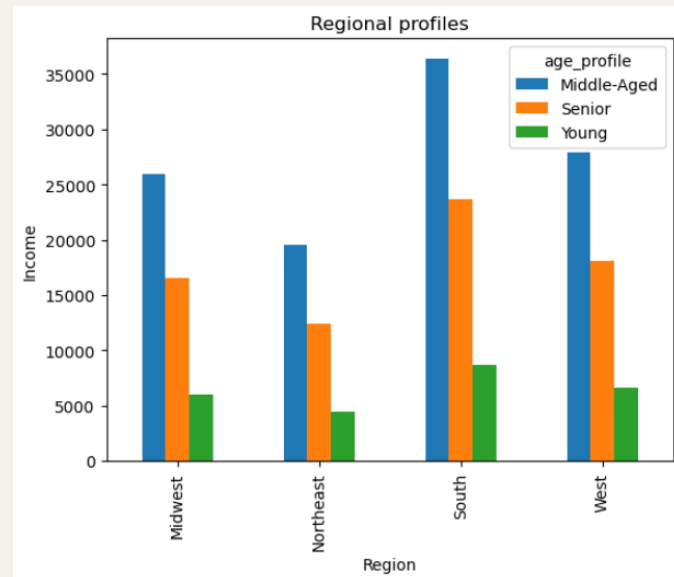


Sunday (0) is the  
busiest day of the week.  
Wednesday (4) is the  
least busy.

# Analysis of Age Profile



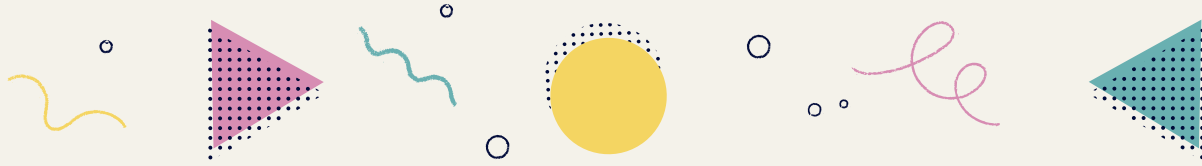
A significant proportion of the customers fall within the middle-aged category, while the senior represents the subsequent largest group. There are very few young customers



The South and West regions have the highest concentration of middle-aged customers, whereas senior customers are evenly distributed across all regions.

# Recommendations

- Tuesday and Wednesday are the days of the week that typically see the lowest volume of orders. Additionally, the morning hours between 07:00-11:00 and evening hours between 18:00-23:00 stand out as potentially high-impact windows for ads.
- As customer profiles become more targeted, less frequented departments should be included if it is shown that the target group prefers items from that department.
- In general, customers tend to spend at similar levels regardless of their income profile, with some minor exceptions. Families tend to spend more on produce compared to single individuals. Additionally, customers aged 40 years or older tend to have a higher income and thus may have more disposable income to spend.



## Additional Sources

Python scripts



Final Report





05

## PIG E. Bank

Identifying customers that  
could depart the bank

# PIG E. Bank

## Project Goal

Providing analytical support to its anti-money-laundering compliance department. contribute to the development and optimization of models that will aid the bank in executing its more efficiently.

## Objective

- The impact of data bias and ethics on how data is used, shared, collected, and protected
- Time-series analysis and time-series forecasting
- Predictive analysis and models such as linear regression

## Data Set

Client Data Set

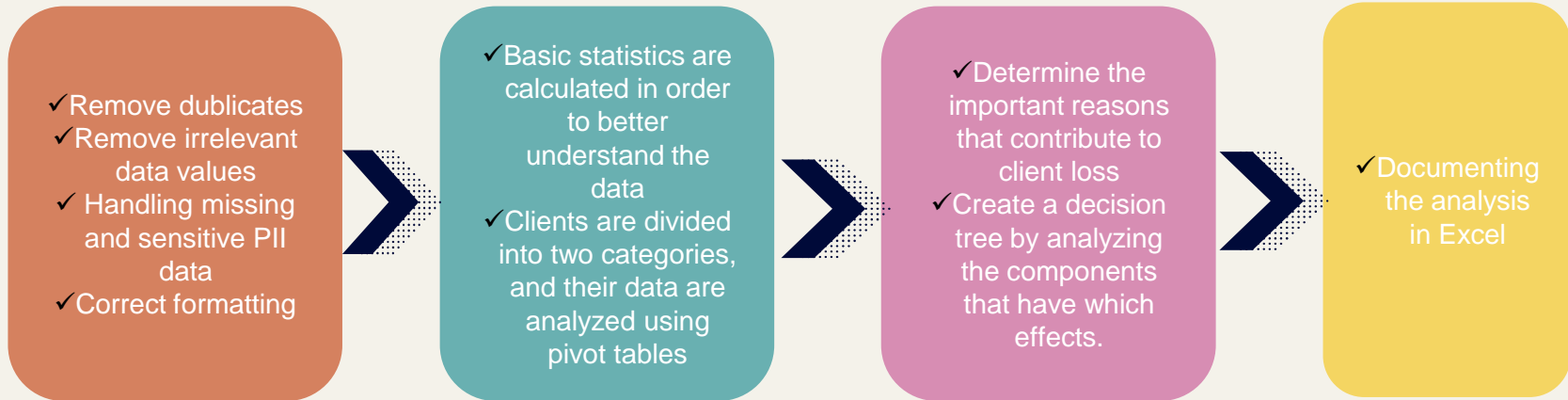
## Tools



## Techniques

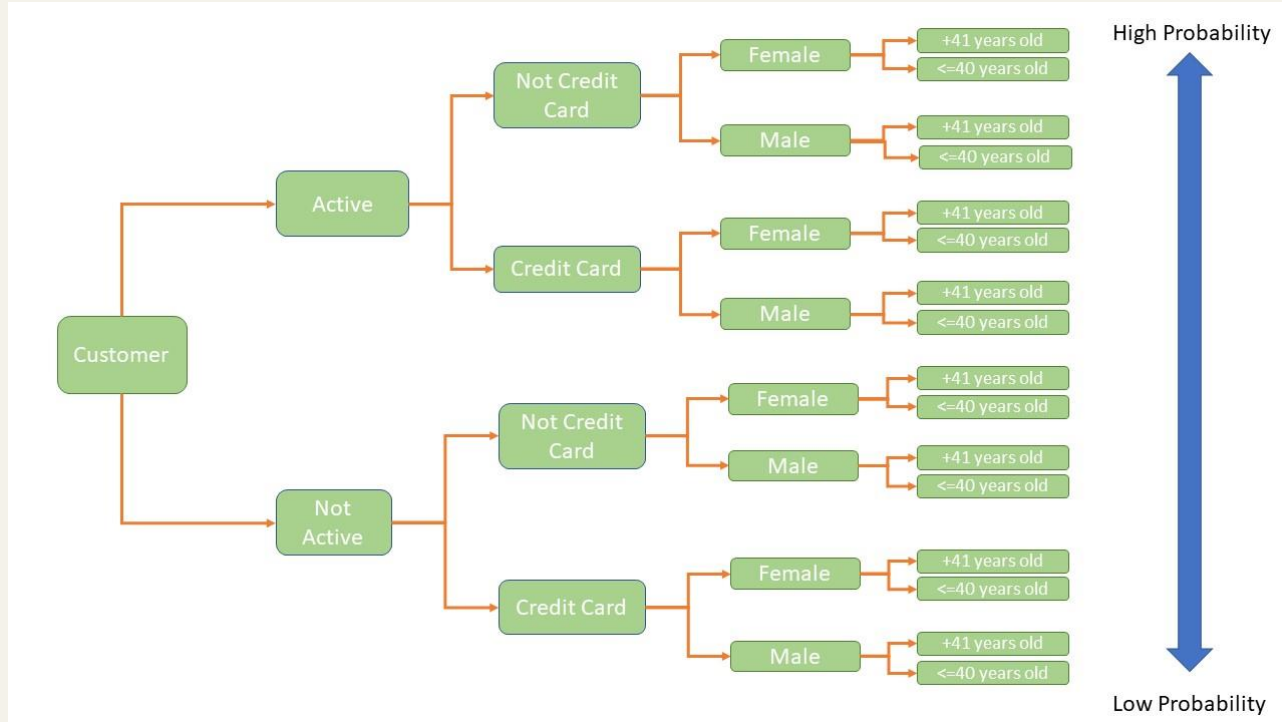
- Data sorting, filtering and cleaning
- Descriptive analysis
- Data ethics
- Data mining
- Building a decision tree as a data mining algorithm

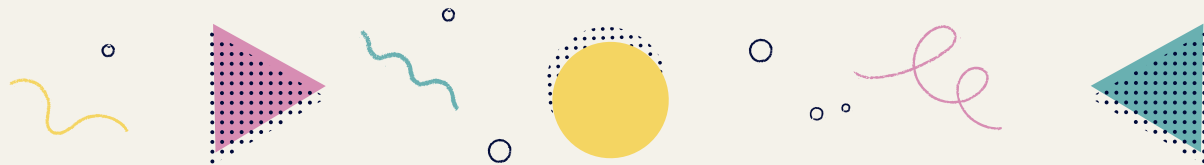
# Data Analysis stages





# Decision Tree





## Additional Sources

Detailed data  
analysis



Decision Tree



Documentation





06

# Global Air Pollution

Contains AQI values of  
different pollutants for many  
cities all over the world.

# Global Air Pollution

## Project Goal

The air pollution is increasing all over the world everyday. The data of the current Air Quality Index (AQI) is ready to analyse the situations of countries.

## Objective

- Determine which countries already having issues
- Determine which countries under risk
- Determine the correlation of AQI with PM2.5, CO, and NO2

## Data Set

Client  
Data Set

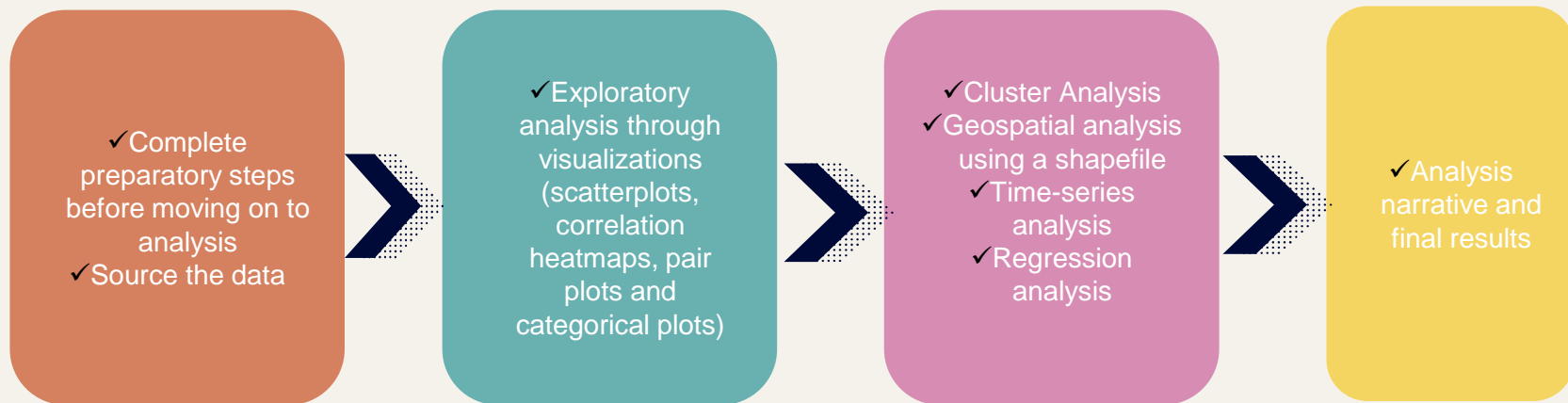
## Tools



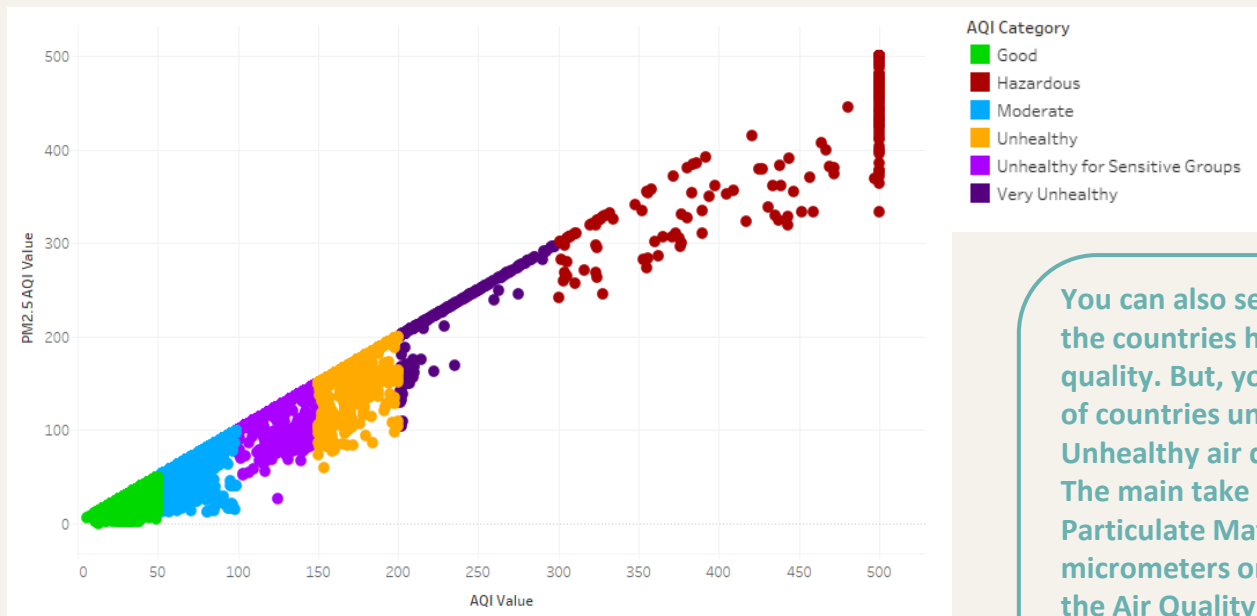
## Techniques

- Data cleaning
- Descriptive analysis
  - Cluster Analysis
  - Spatial analysis
- Time-series analysis
- Regression analysis

# Data Analysis stages



# Exploratory Analysis

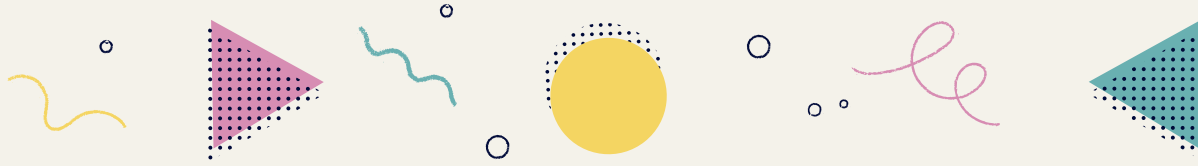


You can also see in this graph that the most of the countries have good or moderate air quality. But, you can also see there are a lot of countries under risk because of having Unhealthy air quality.

The main take away is that, when the Particulate Matter with a diameter of 2.5 micrometers or less (PM2.5) is getting higher, the Air Quality is going bad. The most problematic countries have the most PM2.5 values.

# Summary

- When the Particulate matter (PM2.5) level is increasing in the air, the air quality getting affected very badly based on Exploratory Analysis. Countries must act to reduce particulars to the air.
- The USA, India, Pakistan, Congo and South Africa are the top countries that have bad air quality which is very dangerous for health.
- The most of the countries of Asia, Africa, and South America are under the risk of having Hazardous AQI. The most of the countries of these continents are already having unhealthy air quality.
- The most of the European countries are having good air quality comparing to other continents but France, Italy, Finland are having unhealthy air quality already. Greece, Netherland, Switzerland, Poland are having unhealthy air quality for sensitive groups and they are very close to have unhealthy air quality (Spatial Analysis)



## Additional Sources

Python  
Scripts



Tableau  
Storyboard



Documentation





# Thanks!

Do you have any questions?

sdaonal@hotmail.com



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution