Sherry Tang

Data Analytics Projects

Table of Content

Exploratory Data Analytics

1. SNAP Performance Evaluation (Python, PowerBI)
2. Sports Analysis (R)
3. Coffee Shop Profitability Analysis (R)

SNAP Performance Evaluation & Recommendations

Github link: Data-Analysis-Using-Python/SNAP Performance Measure at main · Sherry-Tang/Data-Analysis-Using-Python (github.com)

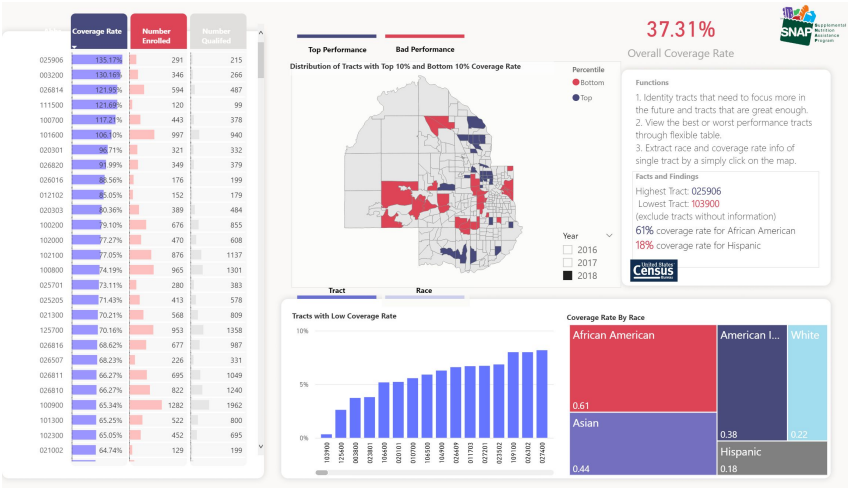
Problem Description: SNAP is a federal program that provides nutrition benefits to supplement the food budget of needy families. Goal of this analysis is to help SNAP understand how does market penetration differ across geography, demographics and time as well as develop actionable strategies.

Analysis Techniques:

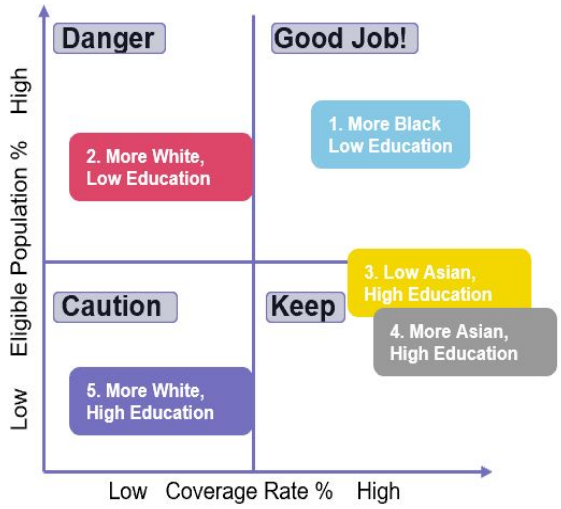
Python, Clustering, Correlation Analysis, Data Visualization

Products:

Power BI dashboard to visualize SNAP coverage rate across time, tract, race and age.



Recommendations:



Groups 1, 3, 4:

- Keep the good performance
- Find out why we're doing well in these tracts

Groups 2, 5:

- Increase reachout personnel for population in **Group 2** (high priority) and **Group 5** (lower priority)
- Find out possible reasons causing the discrepancy in coverage rate
- Work on possibly reducing stigma and negative perceptions in **Group 5**

Sports Analysis

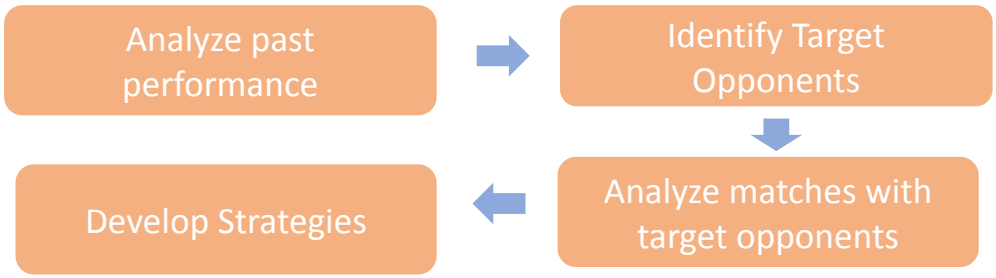
Github link: Data-Analysis-Using-R/Sports Analysis at main · Sherry-Tang/Data-Analysis-Using-R (github.com)

Problem Description: AS Roma is a football team that has consistently been one of the top teams in Italy Series A since 2009. Goal of this analysis is to help the team find patterns that they can exploit to increase success on the field and decrease failure.

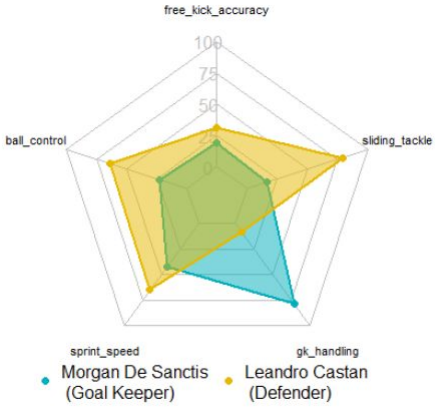
Analysis Techniques:

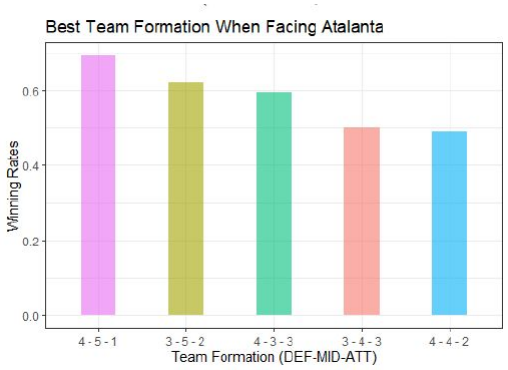
R, Association Rule Mining, Data Visualization

Logic Chain:



Findings & Recommendations:





- ➔ Play more defensively against Atalanta with the team formation of 4-5-1.
- ➔ Play semi defensive with a team formation of 3-4-3 against Inter.
- ➔ Leandro Castan & Morgan De Sanctis together can improve Roma's defense power
- ➔ Francesco Totti & Daniele De Rossi together can improve Roma's attacking power.

Coffee Shop Profitability Analysis

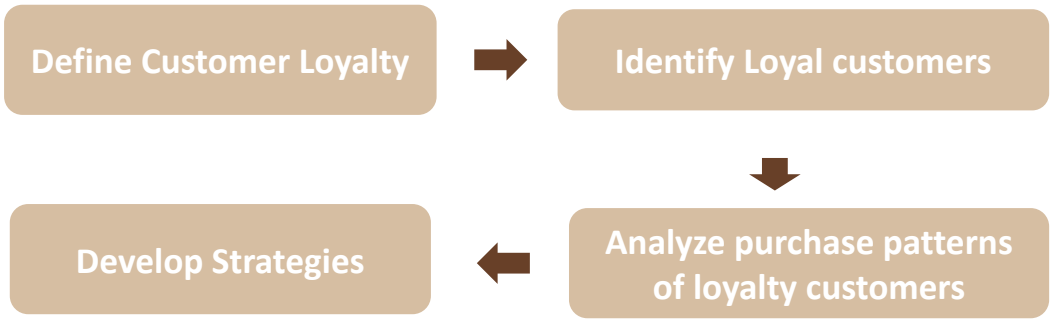
Github link: Data-Analysis-Using-R/Coffee Shop Analysis at main · Sherry-Tang/Data-Analysis-Using-R (github.com)

Problem Description: Goal of this analysis is to help a boutique coffee shop develop strategies to increase customer loyalty and smooth demand.

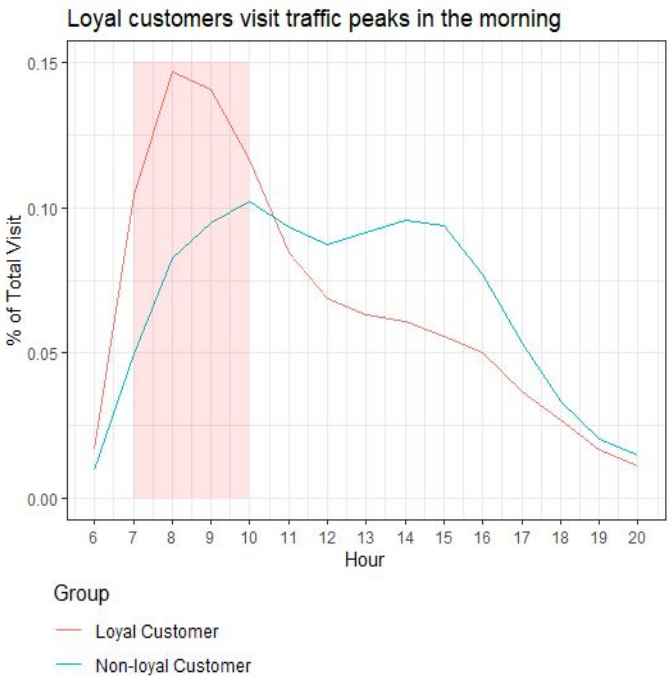
Analysis Techniques:

R, Association Rule Mining, Data Visualization

Logic Chain:



Findings & Recommendations:



Morning:

➔ Pre order service to reduce waiting time

➔ Bundle sale to increase efficiency

Afternoon:

➔ Coupons

➔ Free delivery services

➔ Discount

Machine Learning

1. Santander Customer Transaction Prediction (Python)
2. Microsoft Malware Prediction (Python)
3. Marketing Campaign Conversion Rate Prediction (R)

Santander Customer Transaction Prediction

Github link: Data-Analysis-Using-Python/Standard Transaction Prediction at main · Sherry-Tang/Data-Analysis-Using-Python (github.com)

Project Introduction: Goal of this analysis is to predict which customers will make a transaction in the future

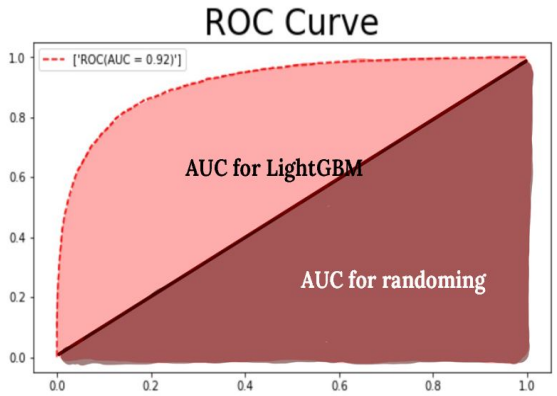
Analysis Techniques:

Python, Predictive Modeling (Light GBM) , Data Visualization

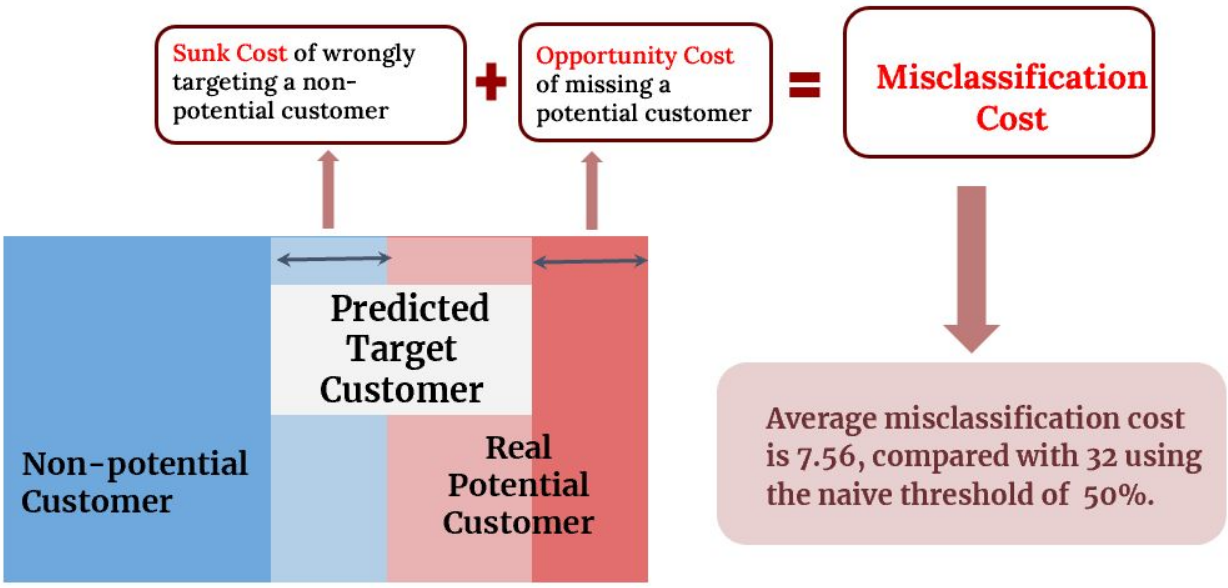
Model Performance:

Best Model →

Model	AUC
LightGBM	0.92282
XGBoost	0.91745
Neural Network	0.88149
Naive Bayes	0.888



Misclassification Cost:



Microsoft Malware Prediction

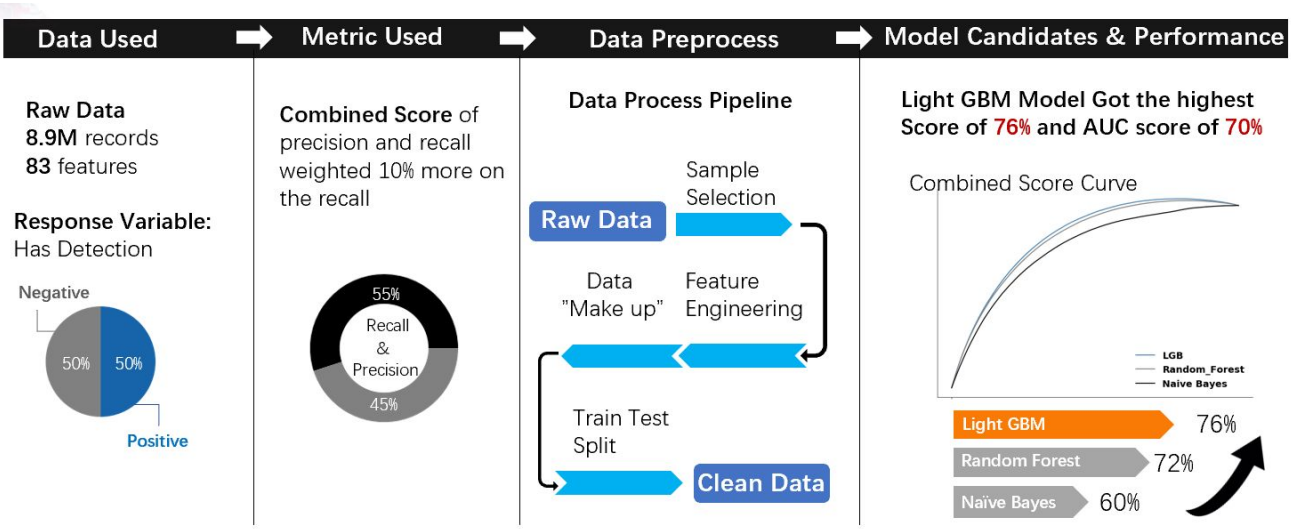
Github link: Data-Analysis-Using-Python/Microsoft Malware Prediction at main · Sherry-Tang/Data-Analysis-Using-Python (github.com)

Project Description: Once a computer is infected by malware, criminals can hurt consumers and enterprises in many ways. Goal of this project is to develop techniques to predict if a machine will soon be hit with malware.

Analysis Techniques:

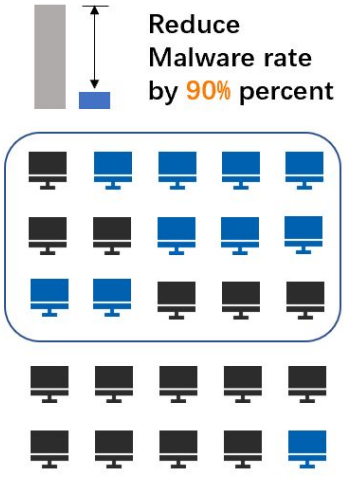
Python, Feature Engineering, Predictive Modeling

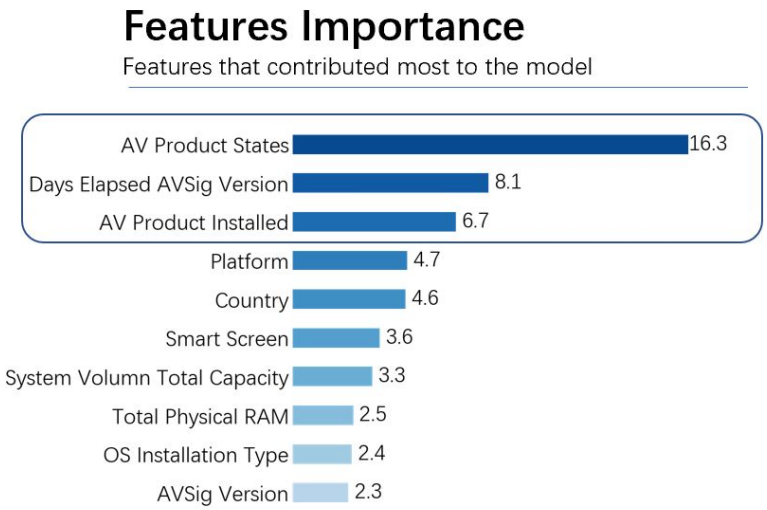
Modeling Process:



Performance Presentation:

Cut off: 0.24 Recall: 90% Precision: 60%





Marketing Campaign Conversion Rate Prediction

Github link: Data-Analysis-Using-R/Marketing Campaign Conversion Forecast at main · Sherry-Tang/Data-Analysis-Using-R (github.com)

Project Description:

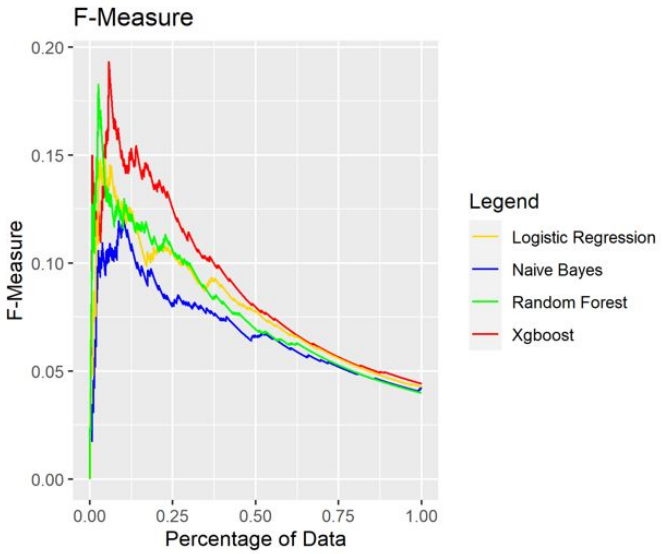
A music-listening social networking website offers basic services for free and provides a number of additional premium capabilities for a monthly subscription fee. Goal of this project is to build the best predictive model for their next marketing campaign.

Analysis Techniques:

R, Feature Engineering, Predictive Modeling (XGBoost)

Model Selection :

- **Logistic Regression**
  - Interoperability Efficient
- **Naive Bayes**
  - Easy Quick Less Accurate
- **Random Forest**
  - Accurate Lack Interpretability
- **Xgboost** ✓
  - Accurate Quick Lack Interpretability



Findings & Recommendations:

Using the model to select target customers can reduce Campaign cost by **82%** and increase adoption rate by **8.15%**

Deep Learning

1. NLP Text Summarization
2. Image Recognition (Python, CNN)
3. Establish Agent Performance Metric (Python)

Establish Agent performance Metric

Due to the confidencial agreement, the project was done on client's virtual machine

Project Description: Calabrio offers a variety of tools for clients to optimize their customer service experiences. Goal of this project is to design a performance metric that can measure agent's ability to build customer rapport.

Analysis Techniques:

Python, Natural Language Processing, Neural Network

Analysis

$$Rapport\ Score = 100 * \frac{Mirroring\ Score + Empathy\ Score + Positivity\ Score}{3}$$

Positivity

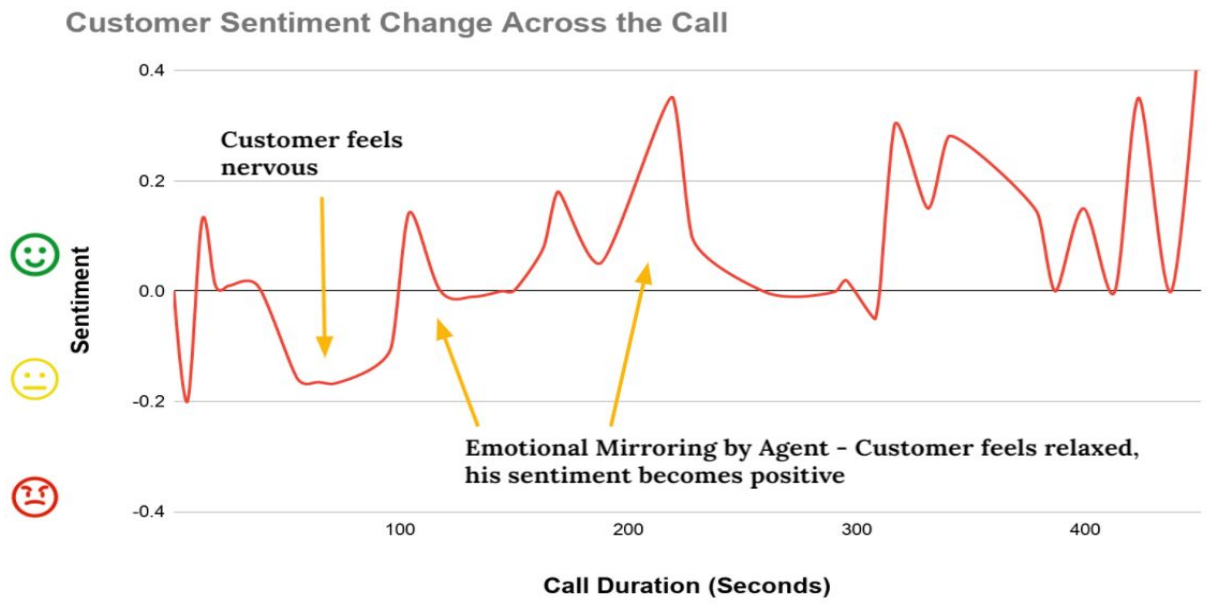
- ➔ Using language that helps the other party feel happier and more at ease.
- ➔ Makes the customer feel comfortable.
- ➔ **Measured by analyzing word choice of agent.**

Mirroring

- ➔ Subtly imitating the communication patterns of the other party.
- ➔ Often shows up with non-verbal cues .
- ➔ **Measured using speech patterns and tone.**

Empathy

- ➔ The ability to connect with the feelings of the customer.
- ➔ The agent gives assurance to the customer.
- ➔ **Measured by assessing usage of certain phrases.**



NLP Text Summarization

Github link: Data-Analysis-Using-Python/Text Summarization-5 Transformer at main · Sherry-Tang/Data-Analysis-Using-Python (github.com)

Project Description: This project use different natural language models to do text summarization of a news article and compare the model performances.

Analysis Techniques:

Python, Natural Language Processing (T5 & NLTK)

T5 Mechanism:

Result Presentation:

T-5 captured more information without sacrifice the conciseness

Hyundai Motor to buy robot maker Boston Dynamics from SoftBank

PUBLISHED FRI, DEC 11 2020•5:00 AM EST

Hyundai Motor Group units and its chairman have agreed to buy an 80% stake in robot maker Boston Dynamics from SoftBank Group Corp, Hyundai Motor Group said on Friday.

Hyundai said the deal values the robot firm at $1.1 billion, suggesting the automaker group offered $880 million for the 80% stake.

Hyundai can leverage robot technology to expand automation at its unionized car factories, as well as design autonomous vehicles like self-driving cars, drones and delivery robots, analysts said.

The new stake comes after newly promoted Hyundai Motor Group chairman Euisun Chung pledged to reduce reliance on traditional car manufacturing, saying robotics would account for 20% of the company's future business, with car-making taking up 50%, followed by urban air mobility at 30%.

Chung will own a 20% stake in Boston Dynamics, while Hyundai Motor and its affiliates, Hyundai Mobis and Hyundai Glovis, will hold a combined 60% stake.

"The transaction will unite capabilities of Hyundai Motor Group and Boston Dynamics to spearhead innovation in future mobility," Chung said in a statement.

KEY POINTS

- Hyundai said the deal values the robot firm at $1.1 billion, suggesting the automaker group offered $880 million for the 80% stake.
- Hyundai can leverage robot technology to expand automation at its unionized car factories, as well as design autonomous vehicles like self-driving cars, drones and delivery robots, analysts said.

Traditional Algorithm:

- Hyundai Motor Group units and its chairman have agreed to buy an 80% stake in robot maker Boston Dynamics from SoftBank Group Corp, Hyundai Motor Group said on Friday.
- "The transaction will unite capabilities of Hyundai Motor Group and Boston Dynamics to spearhead innovation in future mobility," Chung said in a statement

T5 Transformer:

- Hyundai Motor says it has agreed to buy an 80% stake in robot maker Boston Dynamics from SoftBank Group Corp. The deal values the robot firm at $1.1 billion, suggesting Hyundai offered $880 million for the 80% stake.
- Hyundai can leverage robot technology to expand automation at its unionized car factories, as well as design autonomous vehicles

Cat & Dog Image Recognition

Github link: Data-Analysis-Using-Python/Image Classification at main · Sherry-Tang/Data-Analysis-Using-Python (github.com)

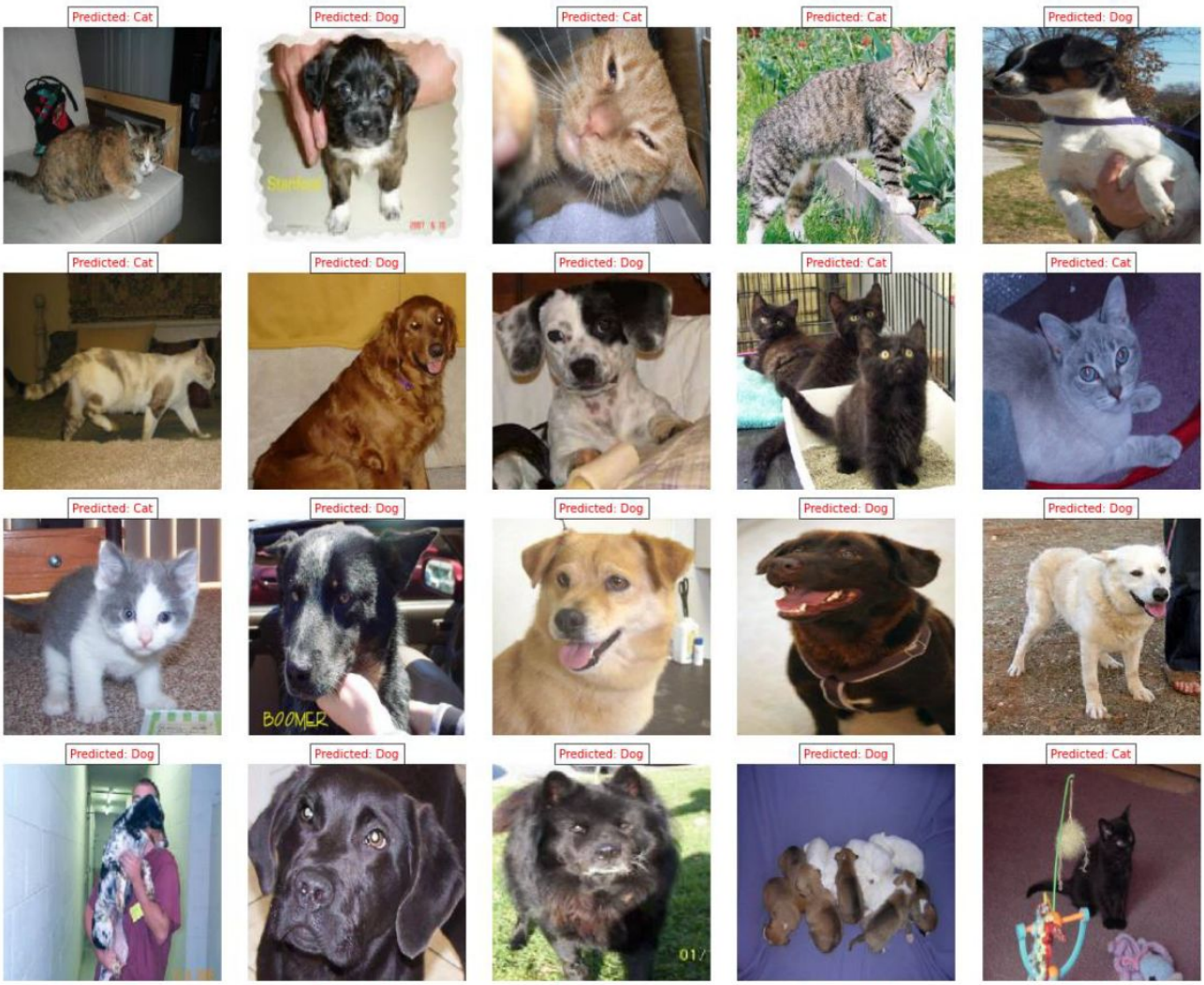
Project Description: Goal of this project is to do the image analysis to identify cats and dogs

Analysis Techniques:

Python, keras CNN with ResNet architecture

Performance Presentation:

Achieved loss score of 0.07 and accuracy of 98%



Office Automation

Auto-Report Design (Python)

Auto Report Design

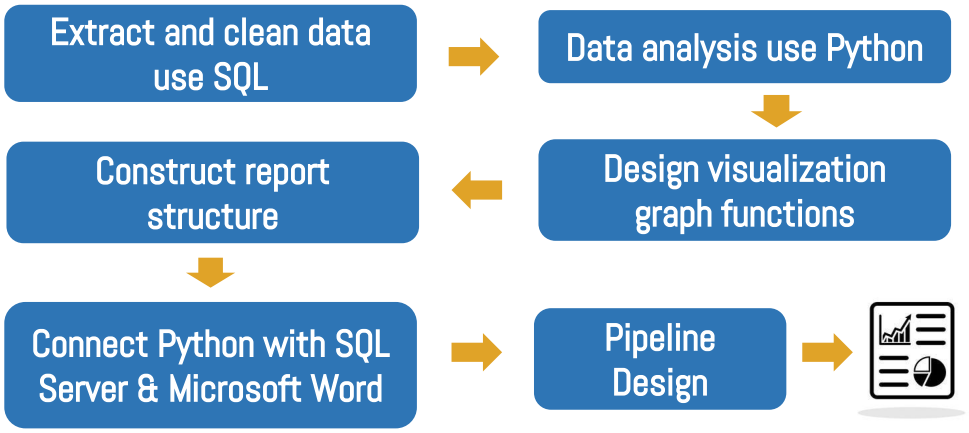
Github link: Data-Analysis-Using-Python/AutoReport at main · Sherry-Tang/Data-Analysis-Using-Python (github.com)

Project Description: This project is to design an automated risk analysis of hoisting machine market.

Analysis Techniques:

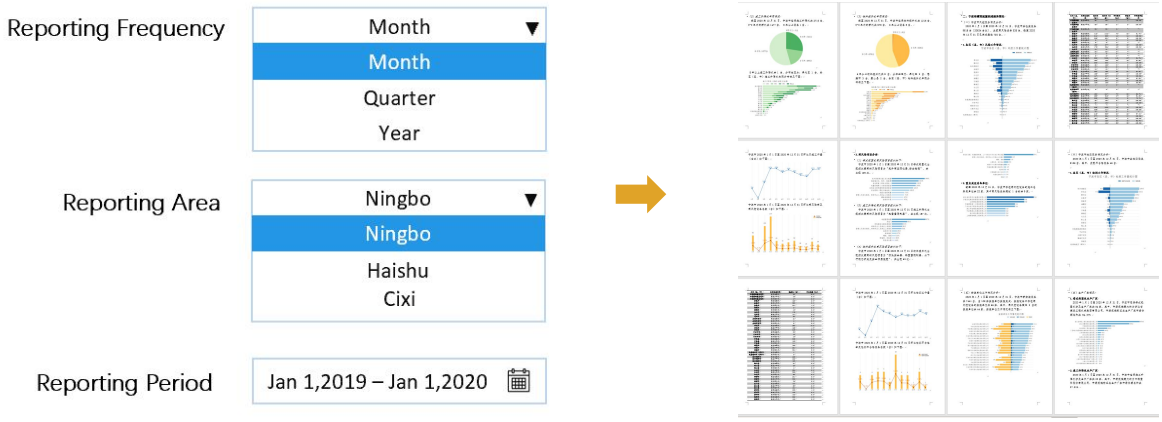
Python , SQL, Data Visualization, Pipeline Design

Process



Result Presentation:

By simply selecting reporting frequency, period and area, the system will automatically generate the analysis report in 30 seconds.



Cloud Computing

Real-Time Customer Sentiment Analysis (AWS)

Real-Time Customer Sentiment Analysis

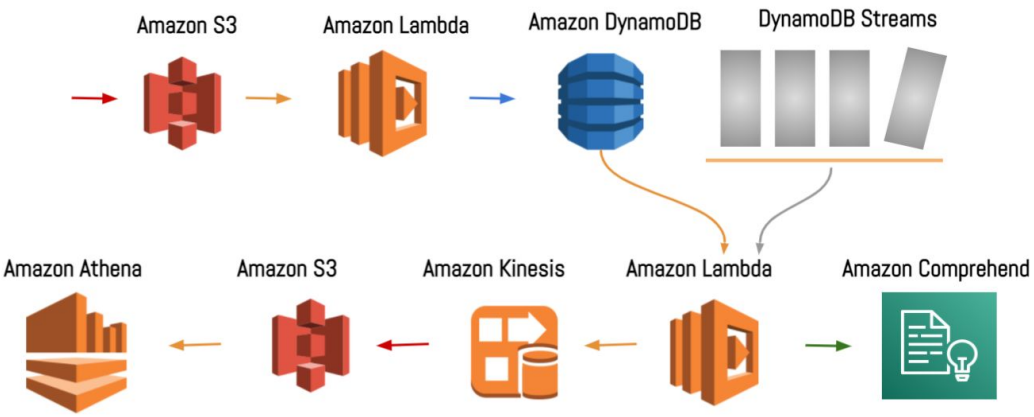
Github link: Big-Data-Analysis/Real-Time Sentiment Analysis (AWS) at main · Sherry-Tang/Big-Data-Analysis (github.com)

Project Description: This project is to design an real-time sentiment analysis pipeline on AWS to provide insights for every single product review and generate real-time summary.

Analysis Techniques:

DynamoDB, Amazon Comprehend

Data Pipeline:



Result Presentation:

- ➔ Vendors could obtain text review data and implement the pipeline in AWS
- ➔ Quickly respond to increased percentage of reviews with negative sentiment
- ➔ Efficient inventory planning for products with growing positive reviews
- ➔ Easily connect to your favorite web apps to share real-time updates