

Restaurant Sales & Status Prediction



【Machine Learning】 【Prediction】 【Algorithms】

July 2021



Chapter 1 - Introduction



- Restaurant owner want to open a new restaurant but don't know anything about the area, district, or other variables
- They will have harder time planning and budgeting than people who has a restaurant based on data and evidence.

Goal: -

I will use restaurant point of sale(POS) data, public data from Hong Kong government, and property agency to determine the best approach for opening a restaurant in Hong Kong.



Benefit : -

- Accelerate the decision-making processes
- We use machine learning to uncover hidden patterns and deliver location insights and recommendations by integrating diverse data sources.

Problem Statement



01

Problem: To get a new restaurant site up and running, a significant amount of time and investment is required.



Outcome: When the wrong location is chosen for a restaurant brand, the site closes within months and incurs operating losses.

02

Problem: Ambros provide over 5000 restaurant data but limited on operational data which may or may not be able to resolve the problem.



Solution: According to number of reference journals, we should add more data such as shop size, location specific data, rental price etc...



Goal: This research is to predict a selection of locations that can survive and the revenue of the restaurants in the given dataset using machine learning algorithms from previously established restaurant data.

Chapter 2 - Literature Review



Point out related researches and then examine the theories that underpin the Problem Statement



Select feature set, target and method algorithms – SVM, KNN, Radom forest regression, ANN, etc...

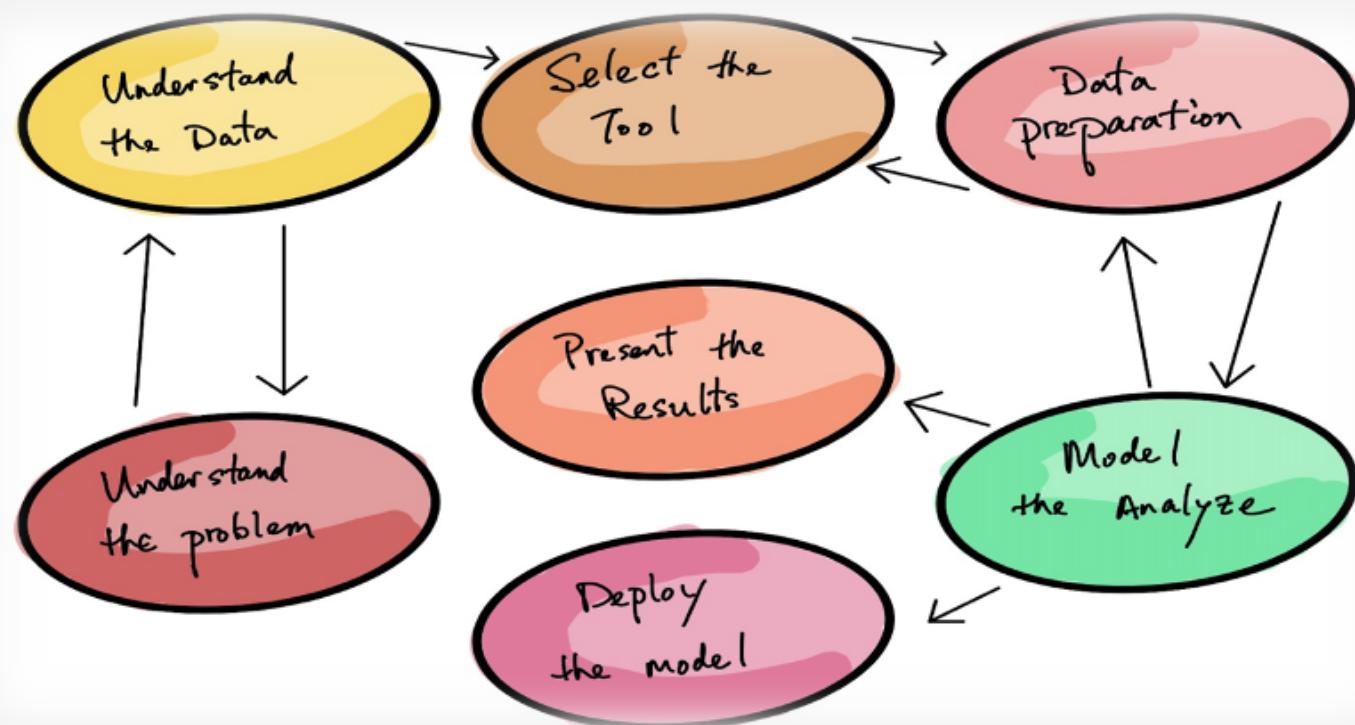
Why Use Machine Learning?

- ML algorithm can simplify code and perform better.
- ML can find complex problems
- ML can easily adapt new data
- Easy to get insights of large amount of data

Chapter 3 - Research Methodology

Define data analysis process:

From Raw data to structured data, model building, evaluation and parameters tunings, visualization and model deployment.



This research of tools in Python

Data Representation



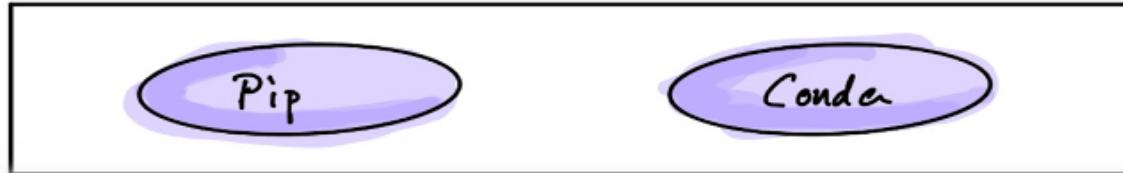
Machine Learning Statistical



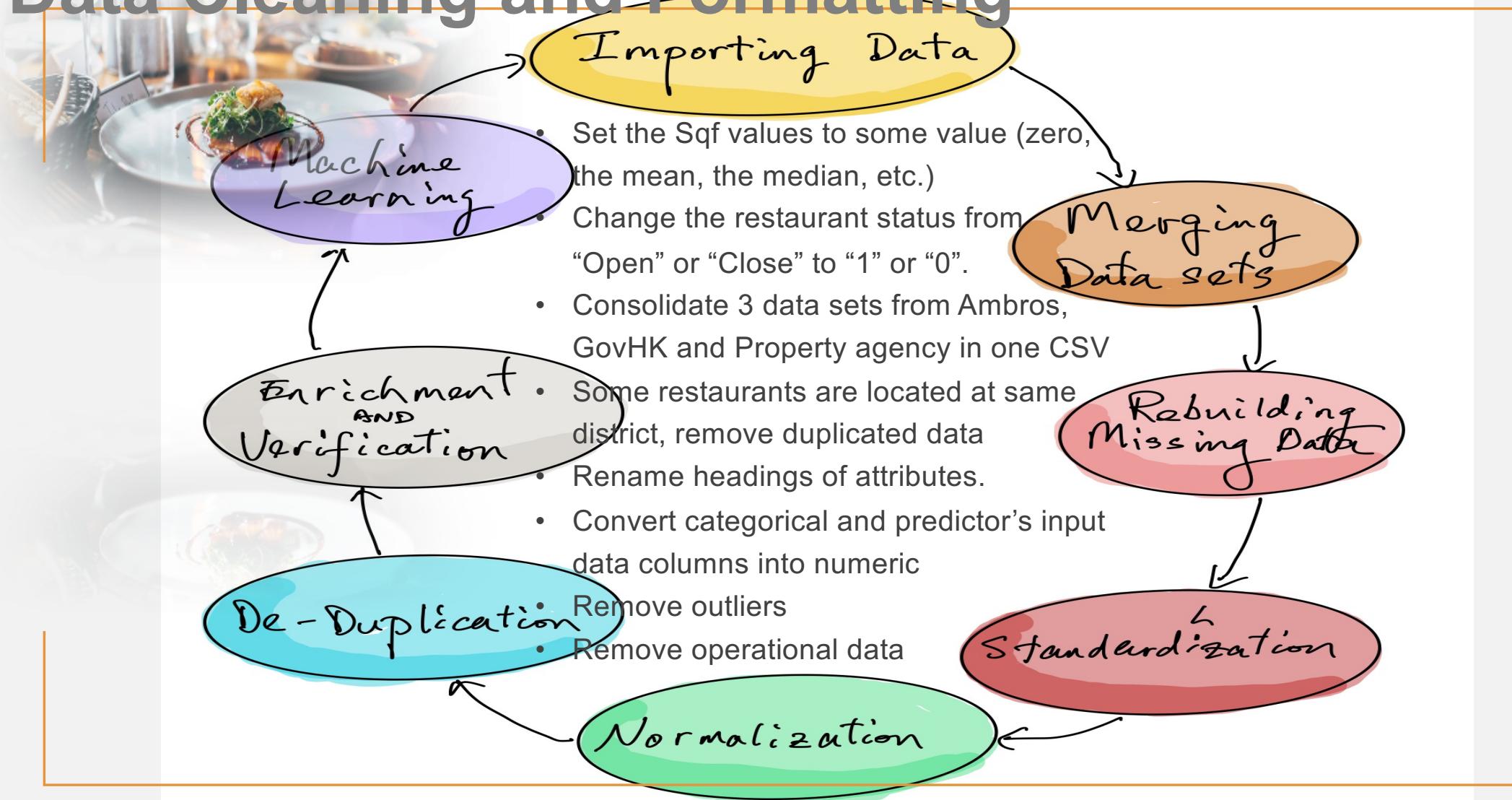
Data Visualisation



Package Management



Data Cleaning and Formatting



Dataset consists of following fields/feature description:



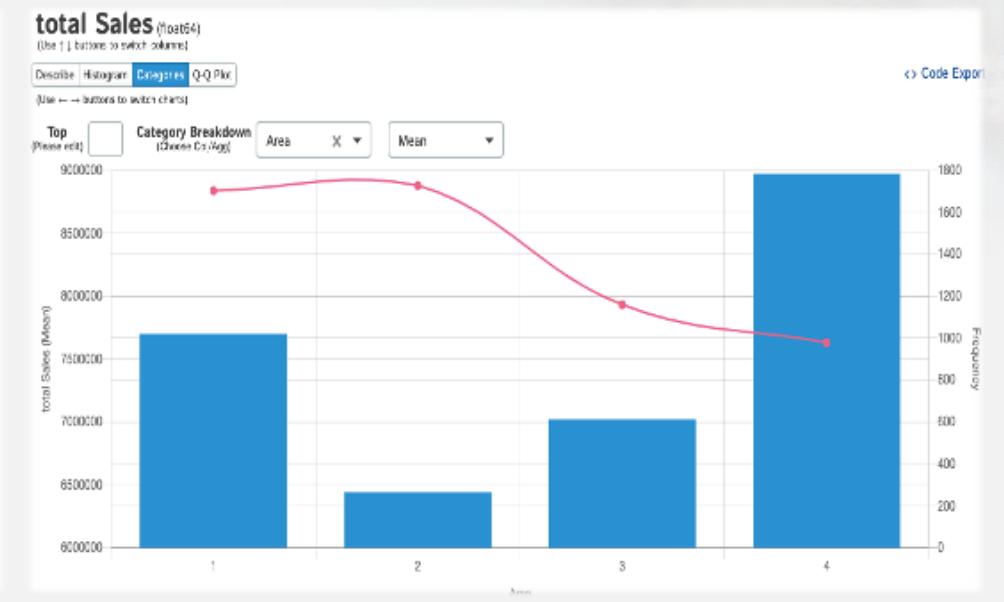
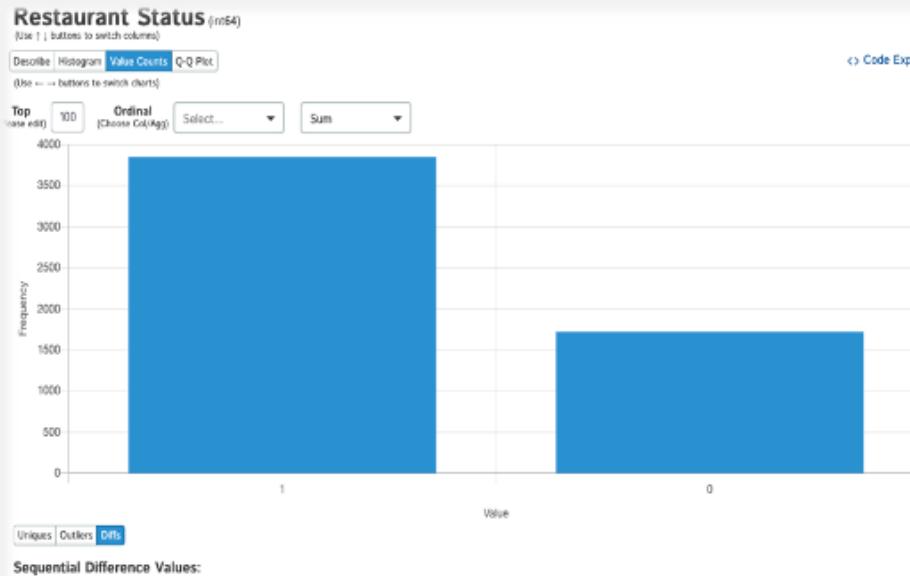
- Restaurant id: the ID of restaurant
- Restaurant type: type of restaurant ie. Chinese, western, fast food, etc..
- Area: restaurant location
- District: HK official district name
- Start Day, month, year
- District about number of people able to work 15 to 65+: Age range is still working in that district
- Area number of worker 15 to 65+: Age range is still working in that location.
- Total Sales: revenue of the restaurant
- Payment type: ie. cash, visa/master, others
- Number of transaction: transaction of the restaurant
- Restaurant Sqf: size of restaurant
- Monthly rent: rental fees
- Location type: eg. shop in the street, mall, school, etc...
- Restaurant Status: Closed or Open



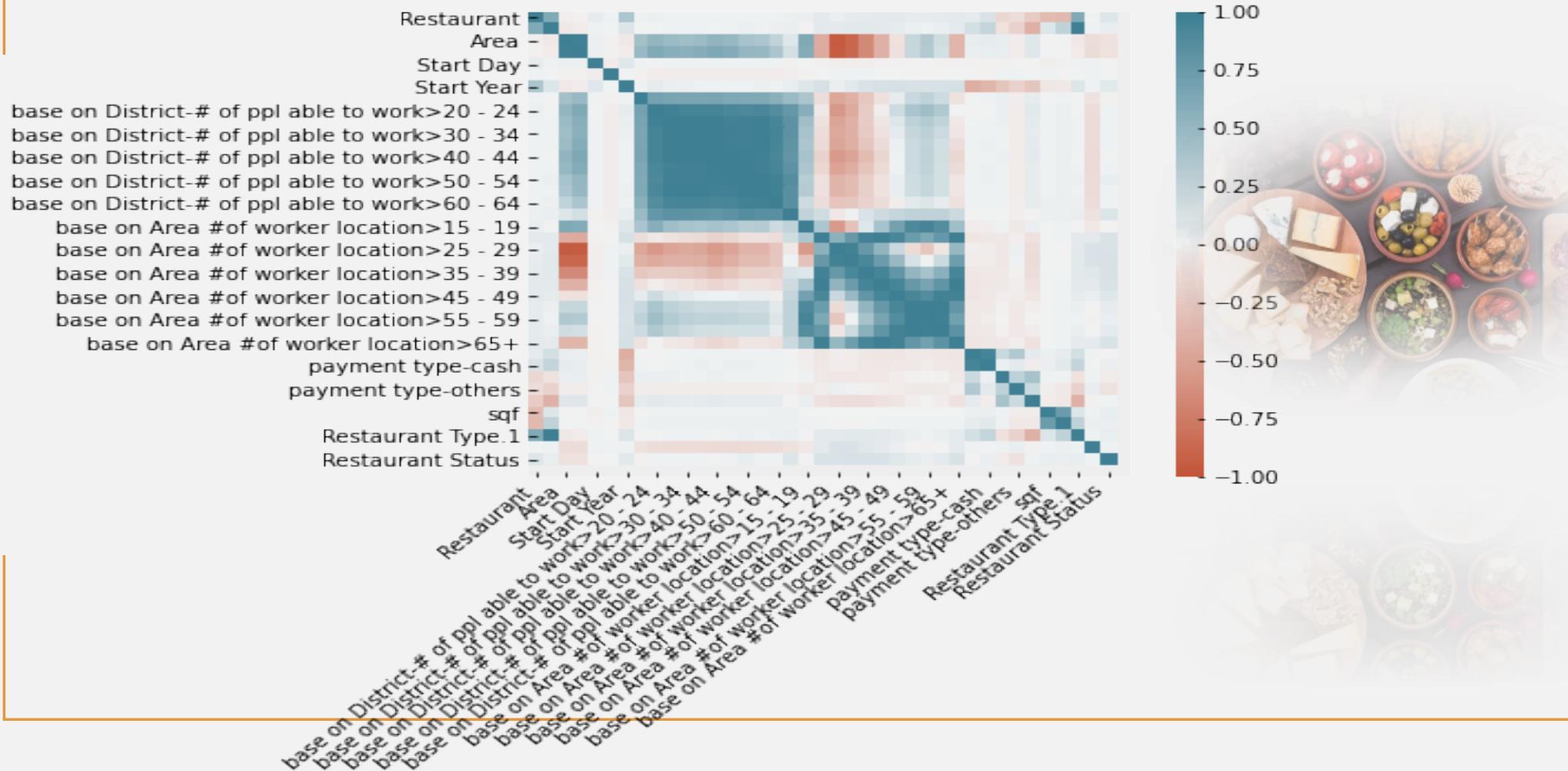
Dataset

Insight of Dataset

- There are 5570 rows and 41 features in training data
- Training data shows variation in the “status” correlation with district who are able to work and “total sales” attribute related to the payment type.
- Area – Kowloon, Kowloon East etc.. District – TST, MK, WC etc.. Type of restaurant – Fast food, Chinese Cuisine, Japanese etc... Location type – Mall, Street, building etc...

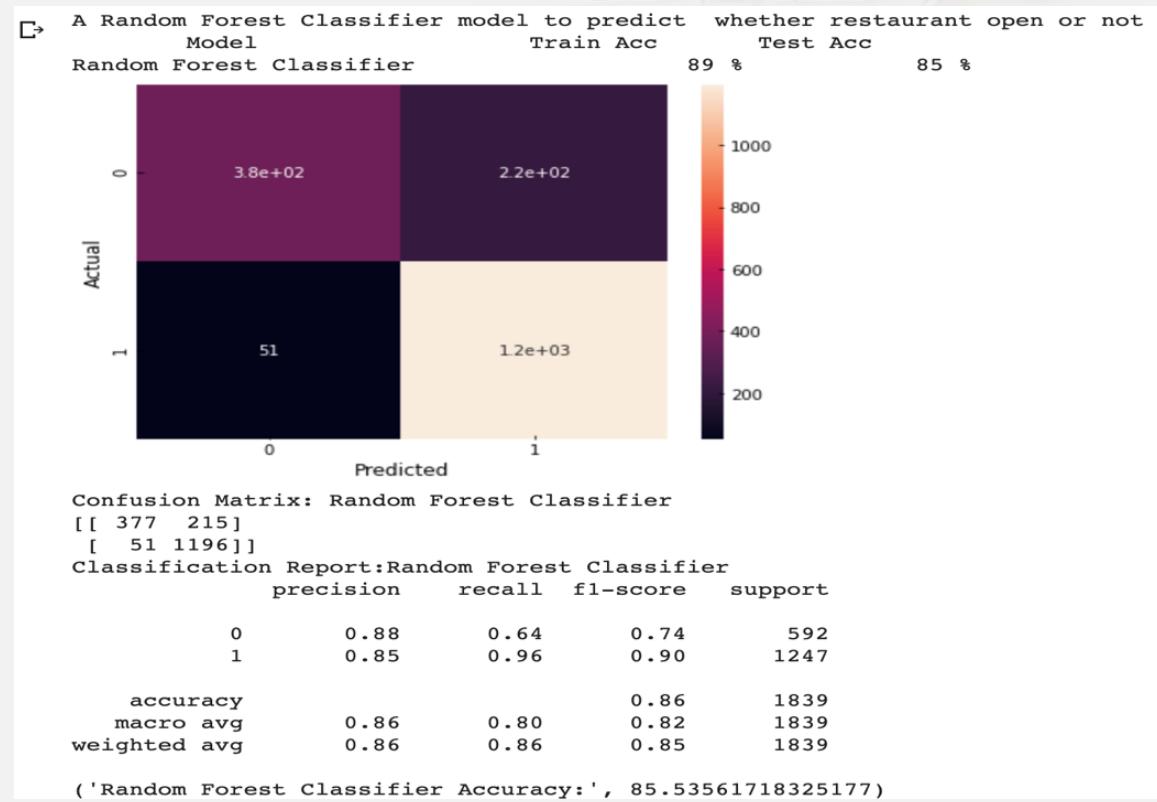
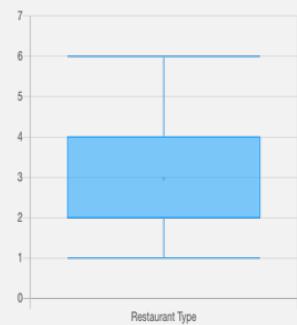
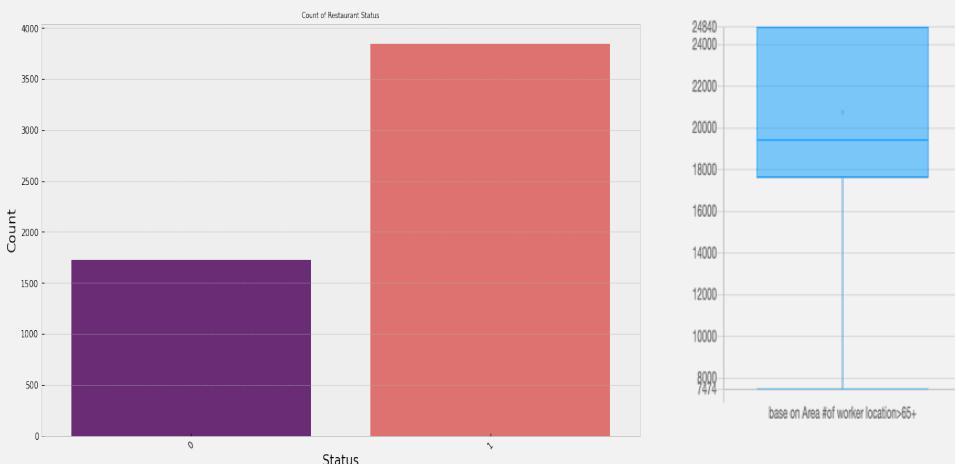


Correlation with all features



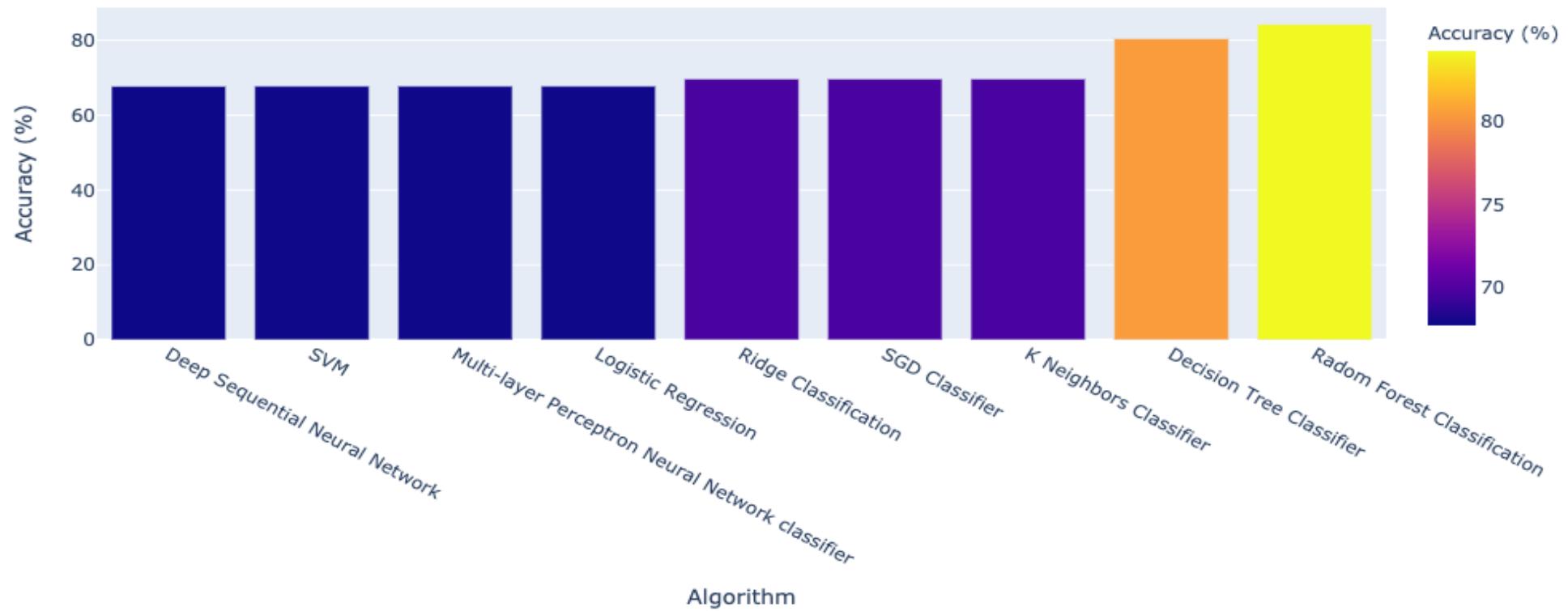
Chapter 4 - Results and Evaluations

Visualization of data, Analysis, Apply ML models and Evaluations

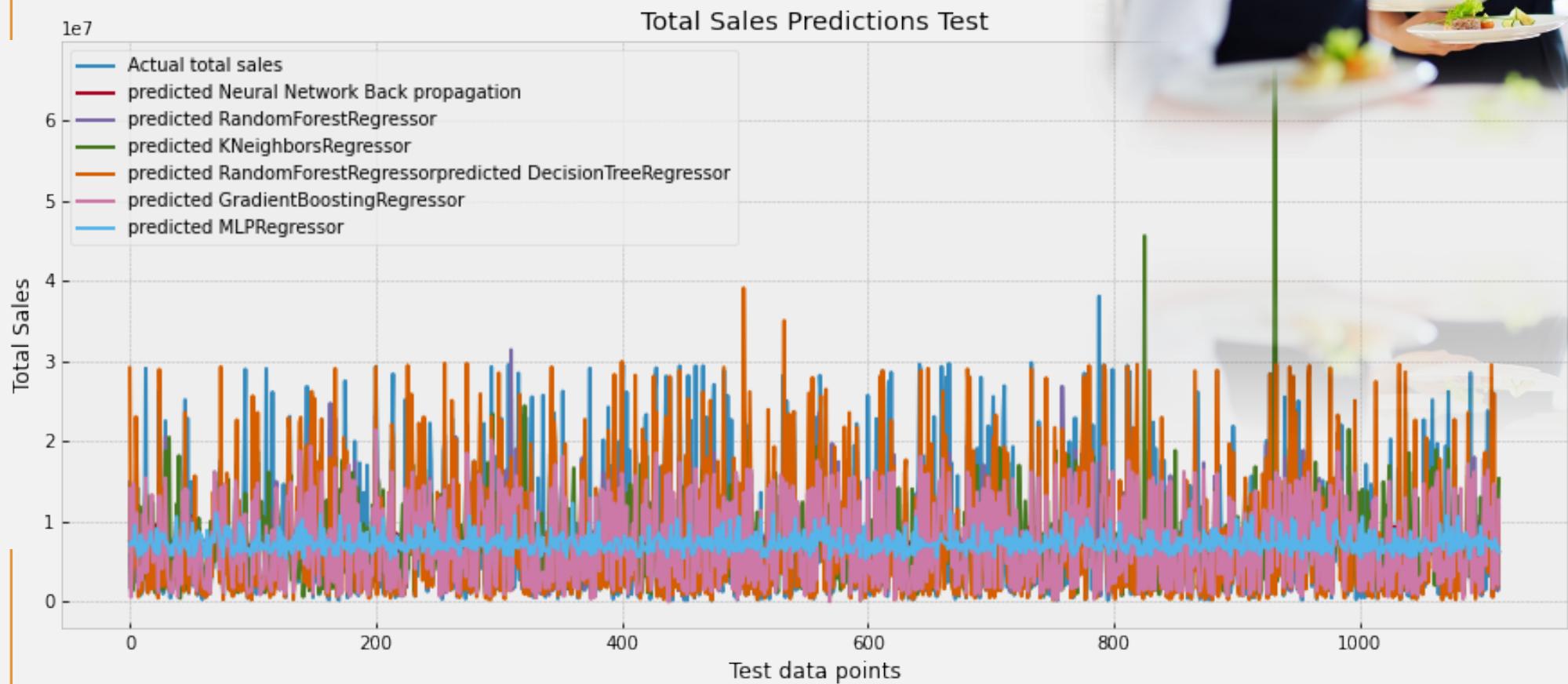


Classifier Forecast Accuracy

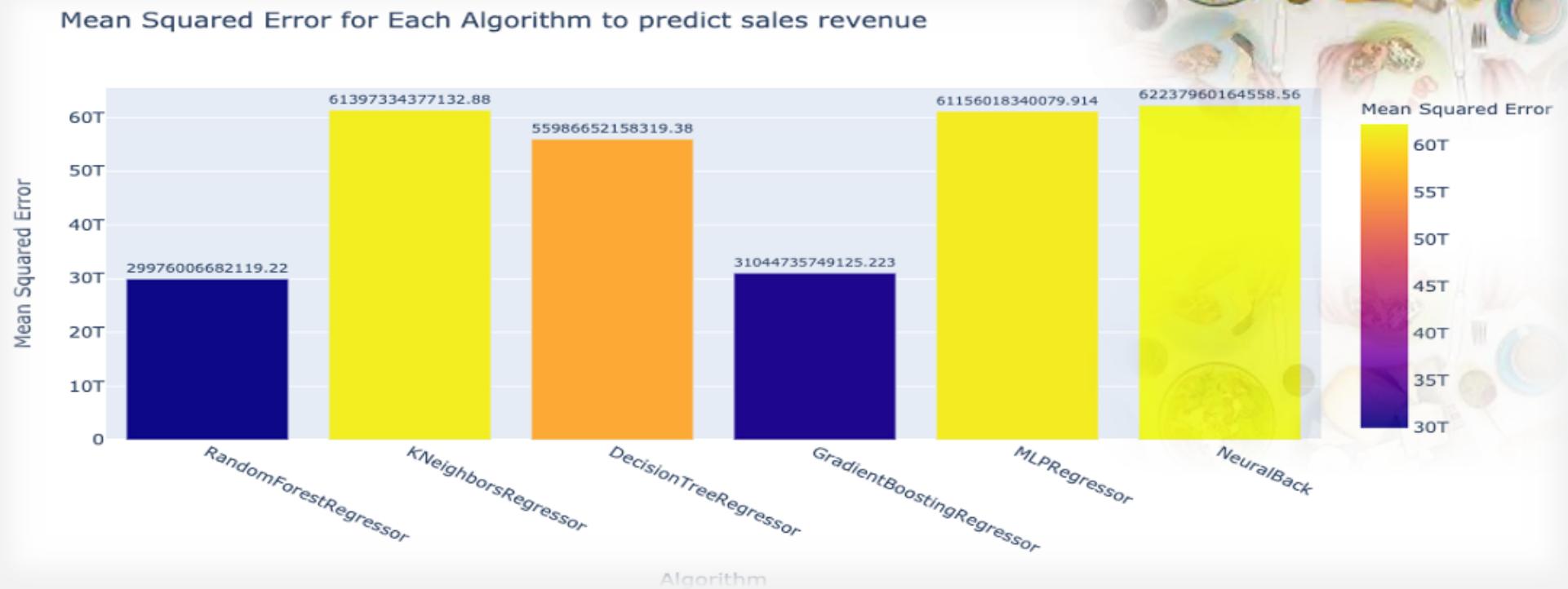
Accuracy of each Classifier for Restaurant Status Recommendation



Total Sales Predictions Test



Mean Squared Error for Each Algorithm to predict total sales



Chapter 5 - Conclusions



- The research shows that the Random Forest Classification is the best method for restaurant's status prediction
- K Neighbour Regression is the best method for restaurant's sales prediction.
- The research found that the machine learning selection must be randomly selected for better accuracy.



End

【Machine Learning】 【Prediction】 【Algorithms】

May 2021

