1. **Clustering (The 3 City)**

Toronto is the capital city of the province of Ontario and the largest city in Canada by population, with 2,731,571 residents in 2016, making it Canada’s most populous city.

Montreal is the most populous municipality in the Canadian province of Quebec and the second-most populous municipality in Canada. Originally called Ville-Marie, or "City of Mary".

Edmonton is the capital city of the Canadian province of Alberta. The city anchors the north end of what Statistics Canada defines as the "Calgary–Edmonton Corridor". The city had a population of 932,546 in 2016, making it Alberta's second-largest city and Canada's fifth-largest municipality.

Given data of point of interests located in 3 cities of Canada (CA) (3City.xlsx), your job is to analyze whether those place is located in Toronto, Edmonton or Montreal.

|  |  |  |
| --- | --- | --- |
| Column | Data Type | Description |
| ID | Integer | The point of interest’s ID |
| Country | Polynomial | Country code of the point of interest |
| PostalCode | Polynomial | The point of interest’s postal code |
| Latitude | Real | The point of interest’s latitude |
| Longitude | Real | The point of interest’s longitude |
| Population | Integer | The point of interest’s population |
| Owner | Polynomial | The point of interest’s owner race |
| Area | Integer | The point of interest’s area |
| Type | Polynomial | The point of interest’s type |
| City | Polynomial | The point of interest’s location (used as **label**) |

1. **Classification (BLUEJEK)**

Bluejek is beyond an app for online transportation, food delivery, logistics, payment, and daily services. By today, Bluejek has partnered with over 1 million drivers, 125.000 merchants, and 30.000 other services, spread across 50 cities in Indonesia.

But not all driver Bluejek hire are giving a good service, complaints about driver services are increasing every day. Human resources division are having a hard time to fire driver one by one.

Given last year’s Bluejek driver data as training data (Bluejek\_Train.xlsx), your job is to predict this year’s driver status whether the driver should be fired or not (Bluejek\_Test.xlsx),

|  |  |  |
| --- | --- | --- |
| Column | Data Type | Description |
| DriverID | Integer | The driver’s ID |
| Name | Polynomial | The driver’s name |
| Gender | Polynomial | The driver’s gender |
| Education | Polynomial | The driver’s education |
| AverageRating | Real | The driver’s average rating |
| BlueFood\_Count | Integer | The driver’s BlueFood transaction count |
| BlueRide\_Count | Integer | The driver’s BlueRide transaction count |
| BlueCar\_Count | Integer | The driver’s BlueCar transaction count |
| WorkingSince | Integer | The driver’s working time (in year) |
| Status | Polynomial | The driver’s status (used as **label**) |

1. **Outlier Detection (Credit Card Fraud Detection)**

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

Given data of transactions made by credit cards in 2017 as training data (Ccfraud\_train.xlsx), your job is to predict which transaction in 2018 is a fraud (Ccfraud\_test.xlsx).

|  |  |  |
| --- | --- | --- |
| Column | Data Type | Description |
| Time | Integer | Number of seconds elapsed between this transaction and the first transaction in the dataset |
| V1…V28 | Real | Principal components obtained with PCA |
| Amount | Real | The transaction’s amount |
| Class | Integer | 1 for fraudulent transactions, 0 otherwise (used as **label**) |