IT8701 –
Introduction to Programming for Data Science

ASSIGNMENT PRESENTATION

13-AUG-2019

Presented By:

Lim Yuan Her (P7359221)

Agenda

- **Dataset**
  - ❑ Source
  - ❑ Nature

- **Design**

- **Analysis Process**
  - ❑ Certificate of Entitlement (COE) Bidding Results
  - ❑ Monthly New Car Registrations by Make
  - ❑ Monthly Motor Vehicle Population by Vehicle Type
  - ❑ Monthly Motor Vehicle Population by Type of Fuel Used

- **Insights**

- **Recommendations and Conclusions**

Dataset - Sources

❑ **Certificate of Entitlement (COE) Bidding Results**
  ▪ https://data.gov.sg/dataset/coe-bidding-results

❑ **Monthly New Car Registrations by Make**
  ▪ https://data.gov.sg/dataset/monthly-new-registration-of-cars-by-make

  ▪ https://www.mytransport.sg/content/dam/datamall/datasets/Facts_Figures/Vehicle%20Registration/Monthly%20New%20Registration%20of%20Cars%20by%20Make.zip

❑ **Monthly Motor Vehicle Population by Vehicle Type**
  ▪ https://data.gov.sg/dataset/monthly-motor-vehicle-population-by-type

  ▪ https://www.mytransport.sg/content/dam/datamall/datasets/Facts_Figures/Vehicle%20Population/Monthly%20Motor%20Vehicle%20Population%20by%20Vehicle%20Type.zip

❑ **Monthly Motor Vehicle Population by Type of Fuel Used**
  ▪ https://data.gov.sg/dataset/monthly-motor-vehicle-population-by-type-of-fuel-used

  ▪ https://www.mytransport.sg/content/dam/datamall/datasets/Facts_Figures/Vehicle%20Population/Monthly%20Motor%20Vehicle%20Population%20Statistics%20by%20Type%20of%20Fuel%20Used.zip

SINGAPORE POLYTECHNIC | SP

Dataset – COE Bidding Results

❑ **Certificate of Entitlement (COE) Bidding Results**

- Dataset consists of 1,120 records with 7 fields as follows:

month	bidding_no	vehicle_class	quota	bids_success	bids_received	premium

- Sample record from the dataset is as follows:

month	bidding_no	vehicle_class	quota	bids_success	bids_received	premium
2010-01	1	Category A	1152	1145	1342	18502
2010-01	1	Category B	687	679	883	19190
2010-01	1	Category C	173	173	265	19001
2010-01	1	Category D	373	365	509	889
2010-01	1	Category E	586	567	1011	19889

- Dataset consists of COE bidding results spanning from January 2010 to April 2019

❑ **Monthly New Car Registrations by Make**

- Dataset consists of 14,160 records with 5 fields as follows:

month	make	fuel_type	vehicle_type	number

- Sample record from the dataset is as follows:

month	make	fuel_type	vehicle_type	number
2016-01	ALFA ROMEO	Petrol	Hatchback	3
2016-01	ALPINA	Petrol	Hatchback	0
2016-01	ASTON MARTIN	Petrol	Hatchback	0
2016-01	AUDI	Petrol	Hatchback	45
2016-01	B.M.W.	Petrol	Hatchback	5

- Dataset contains monthly car registrations by manufacturer e.g. Toyota, Honda etc. from January 2016 to May 2019.

❑ **Monthly Motor Vehicle Population by Vehicle Type**

- Dataset consists of 534 records with 3 fields as follows:

month	vehicle_type	number

- Sample record from the dataset is as follows:

month	vehicle_type	number
2012-01	Cars	593555
2012-01	Rental cars	13970
2012-01	Taxis	27059
2012-01	Buses	17037
2012-01	Goods & other vehicles	159854

- Dataset contains monthly motor vehicle population by type e.g. car, taxi etc. from January 2016 to May 2019.

SINGAPORE POLYTECHNIC | SP

❑ **Monthly Motor Vehicle Population by Type of Fuel Used**

- Dataset consists of 1,152 records with 4 fields as follows:

| month	category	type	number

- Sample record from the dataset is as follows:

month	category	type	number
2016-01	Cars	Petrol	586980
2016-01	Cars	Diesel	6297
2016-01	Cars	Petrol-Electric	6476
2016-01	Cars	Petrol-Electric (Plug-In)	108
2016-01	Cars	Petrol-CNG	1917

- Dataset contains monthly motor vehicle population by fuel ype e.g. petrol, diesel etc. from January 2016 to May 2019.

SINGAPORE POLYTECHNIC | SP

Design

- **<u>Pandas</u>** was used for data manipulation e.g. data cleaning, preparation.

- **<u>Bokeh</u>** was used to provide an interactive dashboard for data visualization with user selectable parameters to perform self-analysis and viewing of trends over time. A tabbed interface is used to separate the different plots displayed.

- **<u>Mongo DB</u>** was used for data storage and retrieval of user defined parameters and analysis results, with corresponding "Save Settings" and "Save Results" buttons available in each tab for these functions.

- To start the application, navigate to the application folder and execute the following command:
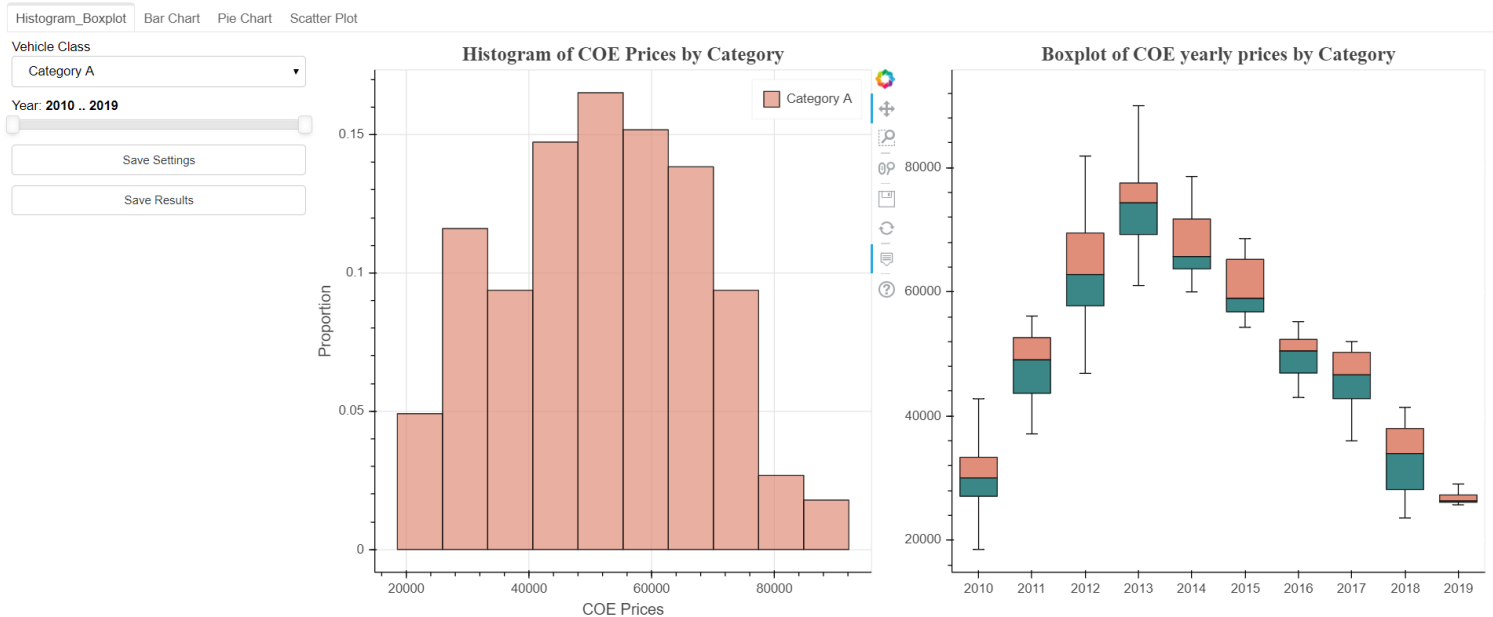
  python app.py

Tabbed interface
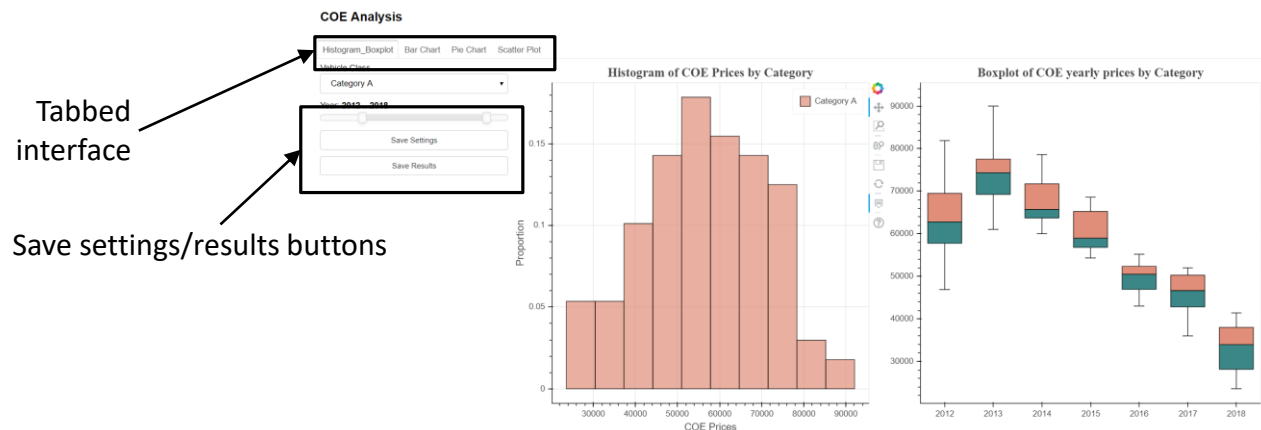
Save settings/results buttons

❑ **COE Premium by period and category**

   ▪ Histograms are used to show the overall distribution of COE bidding premium based on selected period range and category. Boxplots are used to show the yearly premium distribution based on selected period range and category.
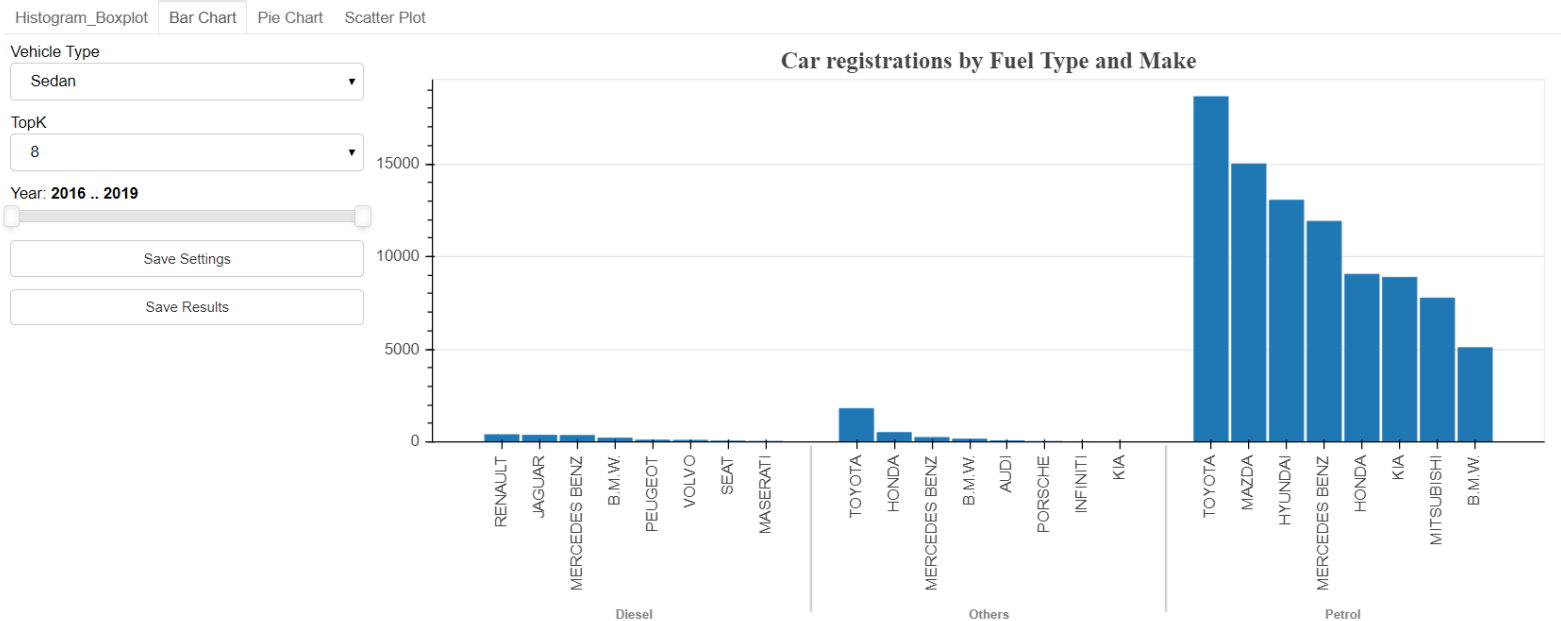
❑ **New cars registered monthly by manufacturer make e.g. Toyota**

▪ Bar charts are used to display the top K manufacturers with highest number of cars registered for selected period range and vehicle type e.g. sedan, SUV etc.

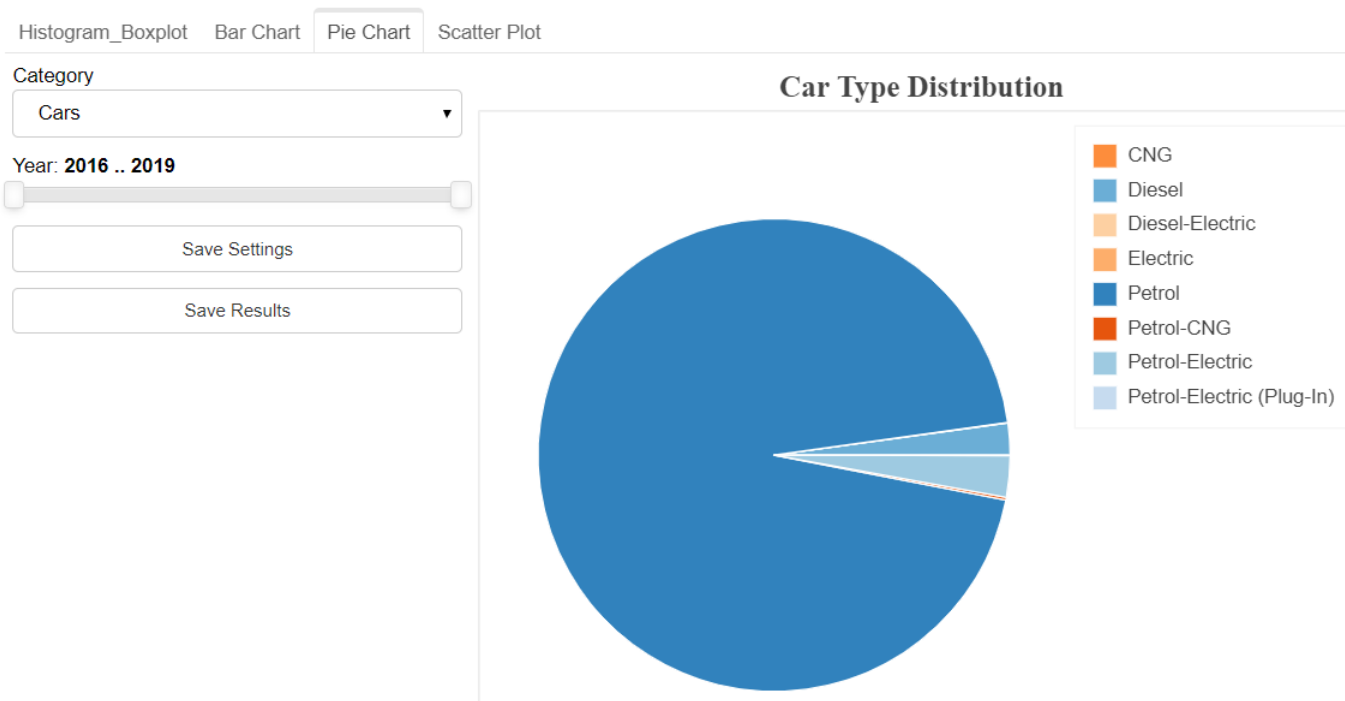▪ Results for each fuel type i.e. Petrol, Diesel, Others are displayed

❑ **Motor Vehicle Population by Vehicle Type**

- Pie charts are used to compare the quantity of vehicles by vehicle type for selected type and period range.

❑ **Motor Vehicle Population by Fuel Type used**

- Scatter plots are used to show the relationship between the number of cars for selected combination of manufacturer make e.g. Toyota, Honda etc. and vehicle type e.g. sedan, SUV etc., and selected vehicle class e.g. Cat A, B etc.
- The top 5 correlation for selected Manfacturer make, vehicle type and vehicle class combination is also displayed. The scatter plot corresponding to the correlation result can be displayed by clicking on the desired result.
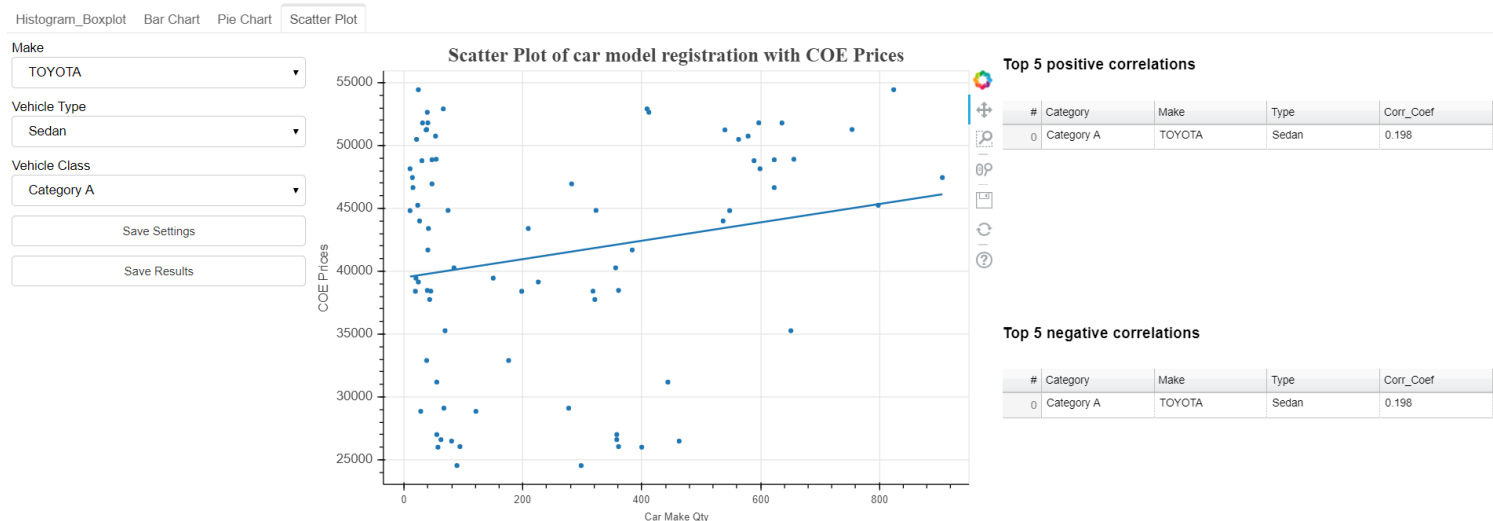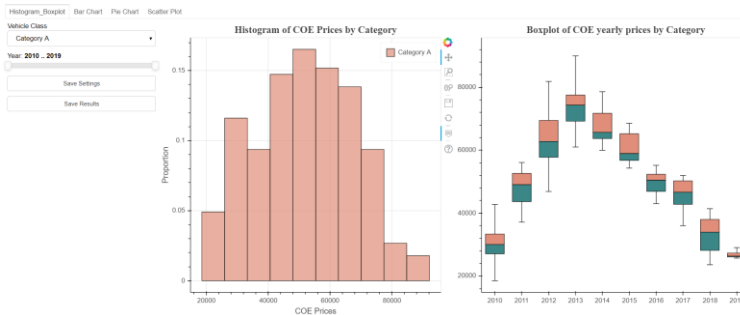
❑ **COE Bidding Results**

  ▪ From the boxplots of yearly COE premiums based on vehicle class, it is observed that **COE premiums peaked in 2012-2013** period for all categories except for Category D motorcycles where premiums peaked in 2018 before reverting below 2014 levels.


Cat A

Cat D

  ▪ From the histograms of COE premium based on vehicle class, it is observed that the premiums are **normally distributed for all categories except for Category D**, which exhibit a bi-normal distribution. The mean COE Premium is approximately **50K** for **Categories A, B, C and E**.

❑ **Monthly New Car Registrations by Make**
 ▪ From the bar charts, the following observations are made
 • **Toyota** leads in the number of vehicles registered for the past 3 years for both Petrol and other fuel types in the **sedan** category.
 • **Mercedes** leads in the **Coupe/Convertible** category for all fuel types.
 • **Honda** leads in the **Hatchback** category for Petrol and other fuel types.
 • **Honda** and **Toyota** leads in the **SUV and MPV** categories for Petrol and other fuel types.
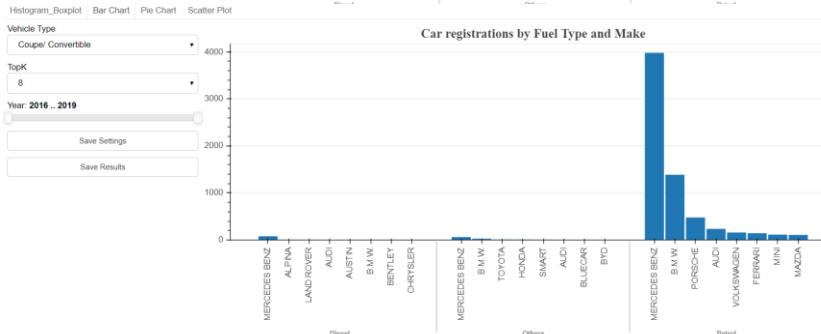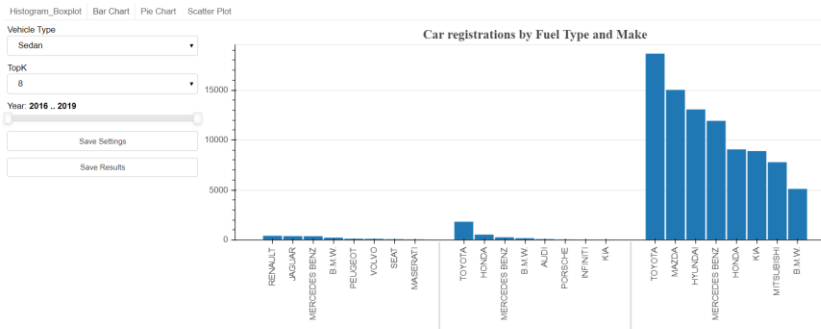
❑ **Monthly Motor Vehicle Population by Vehicle Type**
  ▪ From the pie chart, the following is observed:
    • In the **car** category, increased **growth in hybrid (petrol-electric) vehicles** for the past 4 years
    • In the **Bus and Goods & Other vehicle** categories, **diesel** is still the preferred fuel type used.
    • In the **Taxi** category, majority of taxis still uses diesel, but **hybrid (petrol-electric) taxis have doubled in the past 3 years**.

❑ **Monthly Motor Vehicle Population by Type of Fuel Used**
  ▪ From the correlation results, it is observed that the number of **Mazda SUVs** are **negatively correlated with the COE premiums for Categories A,B and E** i.e. when the COE premiums are low, there are more Mazda SUVs registered.
  ▪ Conversely, **Mazda MPVs and Nissan SUVs** are **positively correlated with the COE premium** i.e. when the OE premiums are higher, more of these cars are registered.

Top 5 positive correlations

#	Category	Make	Type	Corr_Coef
0	Category B	MAZDA	Multi-purpose Vehicle/	0.722
1	Category E	MAZDA	Multi-purpose Vehicle/	0.707
2	Category C	NISSAN	Sports Utility Vehicle	0.703
3	Category A	NISSAN	Sports Utility Vehicle	0.702
4	Category C	MAZDA	Multi-purpose Vehicle/	0.693

Top 5 negative correlations

#	Category	Make	Type	Corr_Coef
0	Category A	MAZDA	Sports Utility Vehicle	-0.801
1	Category A	RENAULT	Multi-purpose Vehicle/	-0.767
2	Category B	MAZDA	Sports Utility Vehicle	-0.76
3	Category E	MAZDA	Sports Utility Vehicle	-0.743
4	Category A	HYUNDAI	Sedan	-0.737

Recommendations and Conclusion

- COE premiums for all categories except Category D have been **declining** since their **peak in 2012-2013** and are at their lowest since 2010. This is perhaps timely for owners looking to replace their existing vehicles.

- The adoption of **hybrid (petrol-electric)** vehicles have been increasing in the past few years, especially in the taxi category where growth have doubled in the past 3 years. Although still a small percentage of private cars, it is slowly increasing in numbers and the trend looks set to continue in the future.

- **Toyota** and **Honda** are the most popular Japanese car brands in most categories whilst **Mercedes** monopolize the **coupe/ convertible** category.

- **Mazda SUVs** are highly sought after when COE premiums are **low** whilst **Mazda MPVs and Nissan SUVs** see higher numbers when COE premiums are **high**.