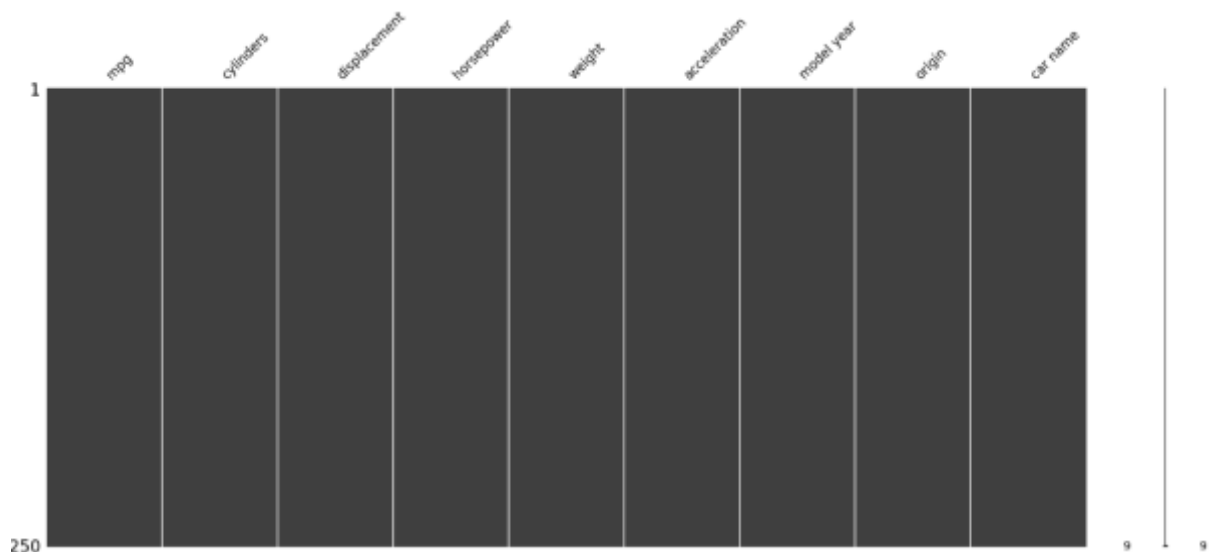


Unit 2 EDA

Several tools from python libraries used to identify missing values, estimate skewness and kurtosis, display correlation as heatmap between variables, scatter plot between correlated variables and encode categorical value with numerical values from auto mpg dataset from kaggle

1. No missing values in the dataset



2. Estimation of skewness and kurtosis

```
(mpg          0.457066
cylinders     0.526922
displacement  0.719645
weight        0.531063
acceleration  0.278777
model year    0.011535
origin        0.923776
dtype: float64,
mpg          -0.510781
cylinders     -1.376662
displacement  -0.746597
weight        -0.785529
acceleration   0.419497
model year    -1.181232
origin        -0.817597
dtype: float64)
```

3. Correlation heatmap

Seminar preparation activities



4. Scatter plot

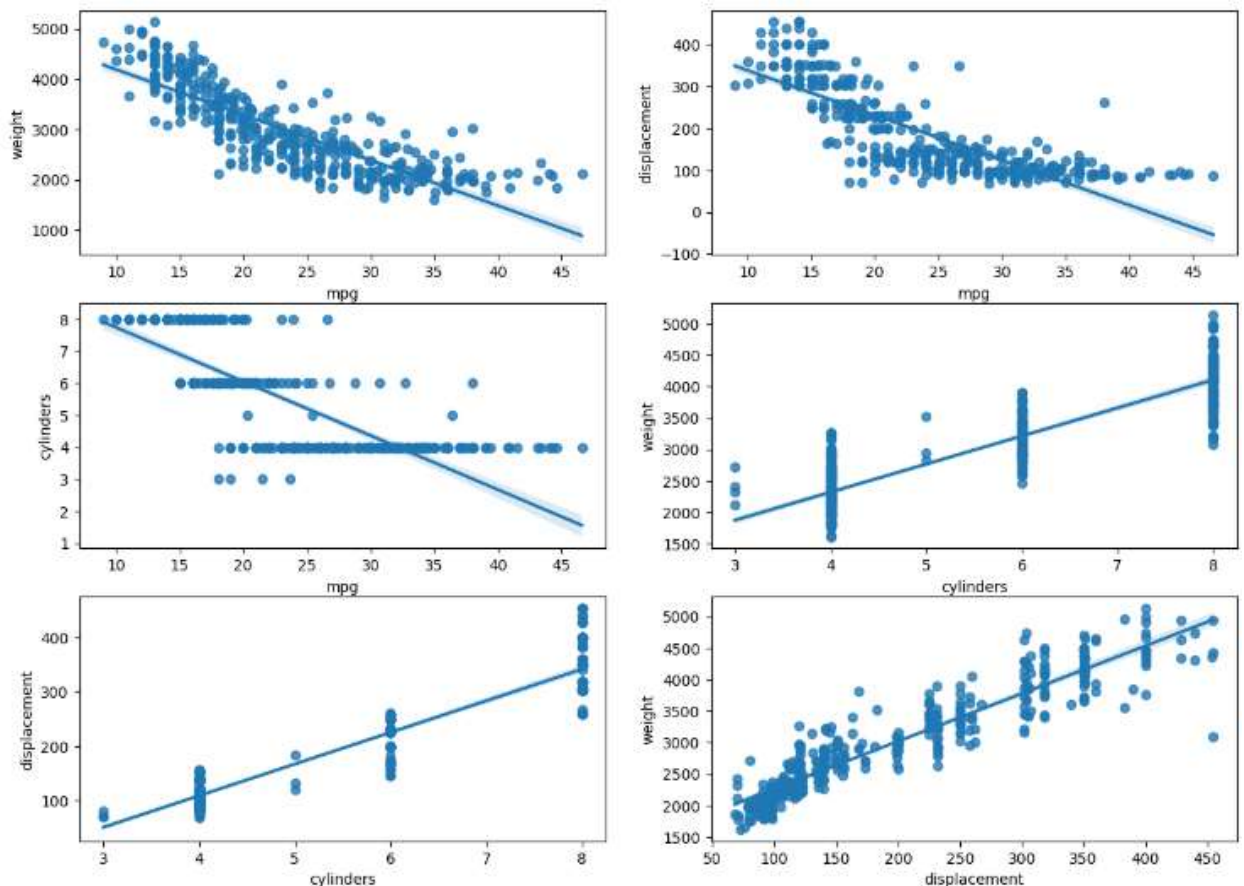
The variables identified to have strong correlation from the heatmap were scatter plotted below

mpg with weight, displacement and cylinders

displacement with weight and cylinders

weight with cylinders

Seminar preparation activities



5. Encoding numerical values to categorical variables

```
from sklearn import preprocessing
#encode label
encode = preprocessing.LabelEncoder()
encode.fit(auto_mpg_data.horsepower)
auto_mpg_data.horsepower=encode.transform(auto_mpg_data.horsepower)

auto_mpg_data.sort_values(by='mpg',ascending=True,inplace=True)
```

Unit 8: Emerging research in ANN

1. Views on the use and impact of ANN

The article '10 Business Applications of Neural Network (With Examples!)' by Mach (2021), describes about the neural network, how they work, types of neurons in the

Seminar preparation activities

neural network and application of neural network in ecommerce, finance, healthcare, security and logistics.

I consider the application of ANN in healthcare valuable in improving the treatment outcome of the patient. In head and neck radiotherapy, ANN could be used to predict the treatment outcome based on correlation of several variables (Cho et al., 2014). While the use of ANN in healthcare looks promising, the availability of quality data for training is questionable. The healthcare community has to setup working groups to collate big data for training ANN, also necessitating good quality data. Also the radiotherapy data can be classified by the Volume, Variety, Velocity and Veracity (4Vs) of Big Data because they are spread across different care providers and not easily shared owing to patient privacy protection (Lustberg, 2017)

Reference:

Mach, P. (2021) 10 Business Applications of Neural Network.

Cho, D.D. et al. (2014) Predicting Radiation Therapy Outcome for Head and Neck Cancer Patients Using Artificial Neural Network (ANN). International Journal of radiation Oncology Biology Physics 90(1): DOI: <https://doi.org/10.1016/j.ijrobp.2014.05.2442>

Lustberg, T., van Soest, J., Jochems, A., Deist, T., van Wijk, Y., Walsh, S., Lambin, P., & Dekker, A. (2017). Big Data in radiation therapy: challenges and opportunities. The British journal of radiology, 90(1069), DOI: <https://doi.org/10.1259/bjr.20160689>

2. Concerns about use of AI in personal insurance.

The Centre for Data Ethics and Innovation has published an independent report on AI and personal insurance on 12 September 2019. This report addresses the use of

Seminar preparation activities

personal data by insurance companies for risk assessments before providing insurance to customers. Queries were raised whether the use of AI for risk assessments would classify some customers as uninsurable. On the other hand, some customers may be classified as trustworthy after using AI. This report also looks at the ethical concerns on data collection and setting up standards on use of data responsibly by the insurers.

As a customer, I feel that the insurance companies should collect only selective details from the customers to issue policies, while sensitive data such as the health records, monthly expense statements should be restricted. Also insurance companies should have policies on how they would handle customer data, to avoid data sharing with their third party stake-holders.

Reference:

CDEI (2019) Snapshot Paper - AI and Personal Insurance.

Unit 10 CNN

I read about CNN in: Wang, J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., Kahng, M. & Chau, P. (2021) CNN Explainer: Learn Convolutional Neural Network (CNN) in your browser. And watched video tutorials

Unit 12: The Future of Machine Learning

The article by Diez-Olivan (2019) mentions that with the current wealth of data, we could effectively describe the behavior of machineries (descriptive), predict the trend of abnormalities (predictive) and setup processes to override damage or critical events (prescriptive), thereby preventing economical losses and ensure safety.

Seminar preparation activities

I have chosen predictive ML model to discuss here, as I consider that as a helpful tool compared to descriptive and prescriptive analysis models. In energy companies, the prediction of energy consumption by customers is very important to meet the demand and supply. Having said that, it is also important to have the customer consumption history for accurate prediction modelling. González-Briones (2019) had reviewed the main machine learning models that allow predicting energy consumption using a one-year data set of a shoe store. 85.7% success was observed for the data set using the Linear Regression and Support Vector Regression models.

Reference:

Diez-Olivan et al. (2019) Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. Information Fusion

González-Briones, A., Hernández, G., Corchado, J.M., Omatu, S., & Mohamad, M.S. (2019). Machine Learning Models for Electricity Consumption Forecasting: A Review. 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 1-6.