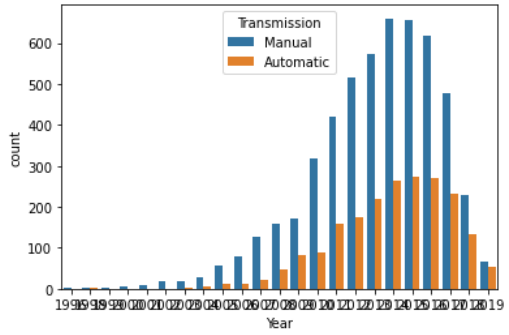
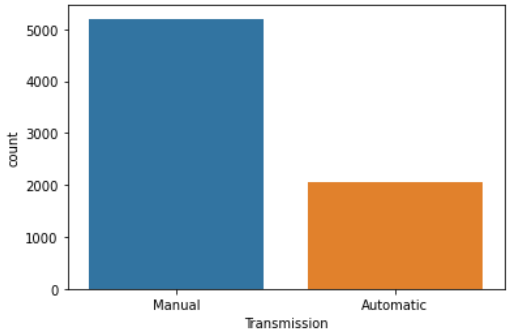
Project 3: Cars4U (Supervised Learning – Foundations)

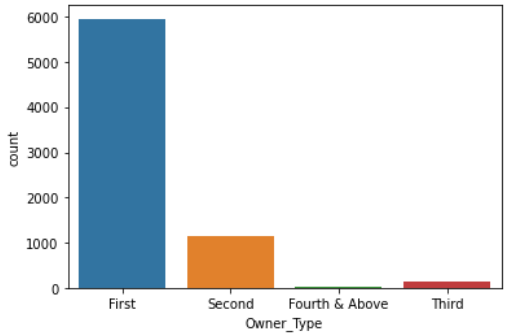
Data Science and Business Analytics
Krithika Srinivasan

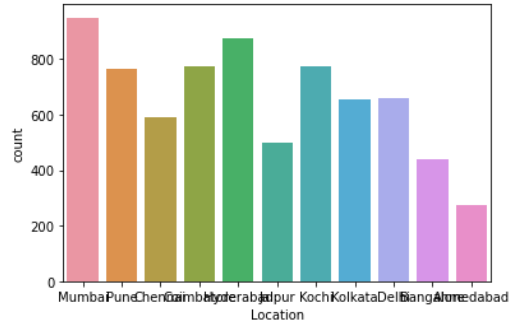
Objectives

- Explore and visualize the dataset.
- Build a linear regression model to predict the prices of used cars.
- Generate a set of insights and recommendations that will help the business.



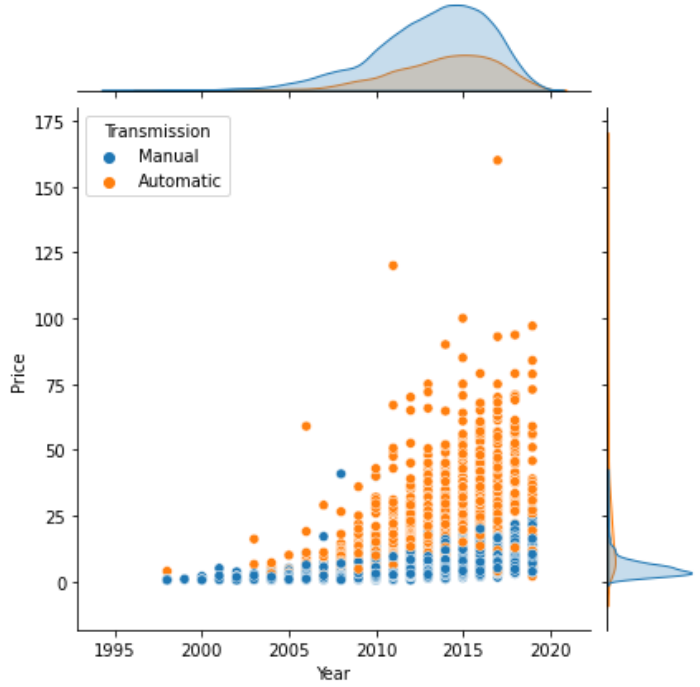


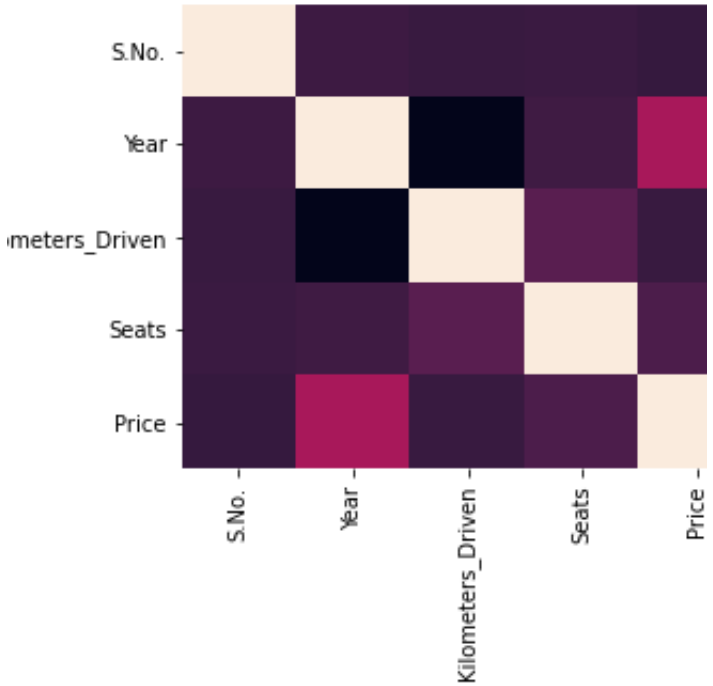


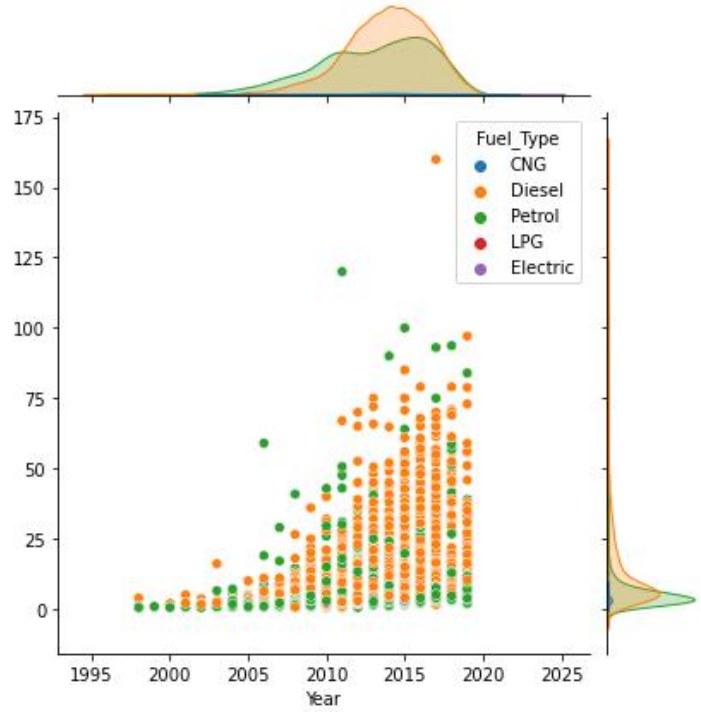


Exploratory Data Analysis (Univariate)

- Average price of used cars sorted by order of most expensive to least expensive location
- Average price of automatic transmission is higher than that of manual
- Electric and Diesel cars are more expensive than other fuel types on average
- Cars that have been owned only by one party are more expensive on average
- Most cars available in Mumbai
- Most cars have been owned by one party







Bivariate Data Analysis

- Some correlation between year and price
- No evidence of strong correlation between any other columns with Price

Data Preprocessing Steps

- Removing suffixes from values in columns- Mileage, Engine, Power, New Price
- Converting New Price to the right units for all values
- Treating missing values
- Encoding categorical columns- Fuel Type, Transmission, Owner Type, Location
- Outlier Detection and Treatment
- Variable Scaling and Feature Selection

Model Building- Linear Regression

- Built the model
- Previous model seemed to have better error metrics
- There is room for improvement – need to fine tune the model through iterations

```
# Checking model performance on train set (seen 70% data)
print("Train Performance\n")
model_perf(linearregression, X_train, y_train)
```

```
Train Performance
```

3]:

	MAE	MAPE	RMSE	R^2
0	0.352425	-37.580729	0.539976	0.719989

```
# Checking model performance on test set (unseen 30% data)
print("Test Performance\n")
model_perf(linearregression, X_test, y_test)
```

```
Test Performance
```

4]:

	MAE	MAPE	RMSE	R^2
0	0.364676	-48.208409	0.532996	0.685448

Recommendations

- Having some better data quality might help make the model prediction better (Example: New_Price column)
- Could process the Name column to determine model of the car
- Acquire more automatic cars as their price is higher
- Acquire more diesel fuel type cars as their price is higher