STUDENT DETAILS

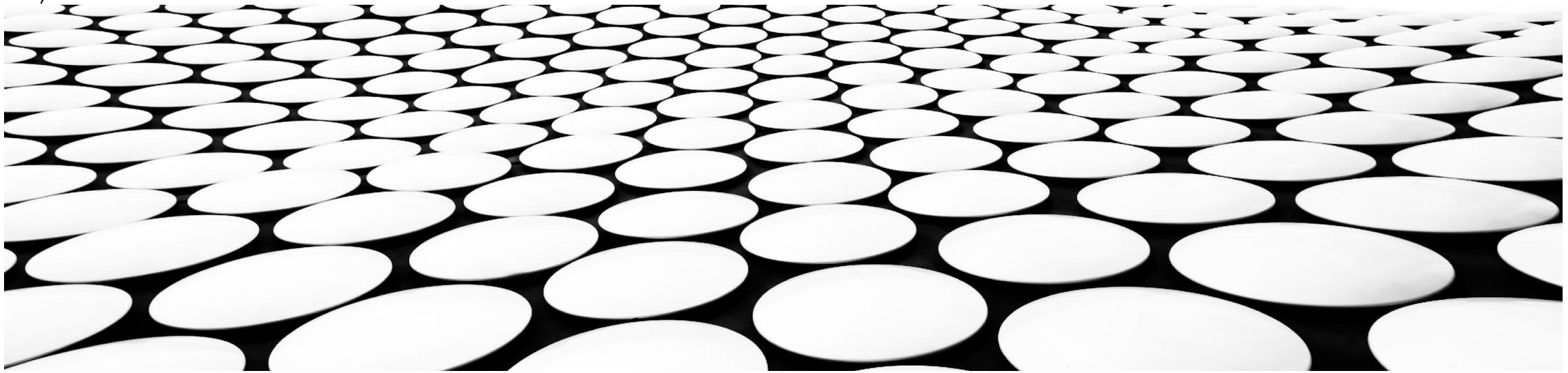
Name: Tanmayee Inaganti

SkillsBuild Email ID: tanmayee_inaganti@srmap.edu.in

College Name: SRM UNIVERSITY AP

College State: Andhra Pradesh

Internship Domain and Internship Start and End Date: Data Analytics / June 3rd, 2024 - July 25, 2024



CAR DEKHO DATA ANALYSIS

The Car Dekho Data Analysis project looks at vehicle sales data to find out what affects prices and sales trends. The project aims to identify patterns and great deals by analyzing details like the year of manufacture, selling price, kilometers driven, and transmission. The goal is to provide insights that help improve sales strategies and understand the market better.

AGENDA

- Problem Statement
- Project Overview
- End Users of the project
- Solution and its Value Proposition
- Customization
- Modelling
- Results
- Links

PROJECT OVERVIEW

The Car Dekho Data Analysis project is an analysis of vehicle sales data to find important trends and patterns. The project uses Python to examine the manufacturing years, selling prices, and variety of vehicles in the data. We can check for any missing information and identify the most sold vehicles. We can also analyze how factors like age and mileage affect prices. The analysis of this data will help to provide insights that can help improve sales strategies.

WHO ARE THE END USERS OF THIS PROJECT?

The following is a list of companies which sell cars, including second hand cars:

- Car Dekho
- CarWale
- Droom
- Spinny
- Cars24

YOUR SOLUTION AND ITS VALUE PROPOSITION

- #From which manufacturing year to which manufacturing year vehicles are present in this data ?
  - min year: 2003, max year: 2018
- #What is the lowest price to which a vehicle is sold ?
  - Lowest price which vehicle is sold: 0.1 lakhs
- #What is the highest price to which a vehicle is sold ?
  - Highest price which vehicle is sold: 35.0 lakhs
- #How many records are there in this data ?
  - 301 records
- #Are there any missing records in this data ?
  - 0 missing records
- #How many different vehicles are present in this data ?
  - 98 different vehicles
- #Which is the most sold vehicle in this data ?
  - city is the most sold vehicle

YOUR SOLUTION AND ITS VALUE PROPOSITION

- #Does the database include any CNG vehicle ? If yes how many of them are there ?
  - 2 cng vehicles
- #How many vehicles here are for sale from Individuals directly ?
  - 106 vehicles for sale from Individuals
- #Does this database contain auto transmission vehicles ? If yes how many of them are there ?
  - 40 auto transmission vehicles
- #How many single person owned vehicles are there in this database ?
  - 290 single person owned vehicles
- #Which is the most and least cost depreciated vehicle in data ?
  - land cruiser is the most cost depreciated vehicle, Honda Activa 4G is the least cost depreciated vehicle
- #Which brands of vehicles are less affected by cost depreciation ?
  - ['Honda Activa 4G'] are less affected by cost depreciation

YOUR SOLUTION AND ITS VALUE PROPOSITION

- #Are there any factors which you feel affect the cost depreciation ?
  - The factors can be Year, Kms Driven, Depreciation to affect the cost depreciation
- #In general selling price is affected by age of vehicle and distance driven by vehicle , is it observable from data ?
  - The answer to this question can be observed from a visualization to observe the data more accurately
- #Can we get idea about newest vehicles i.e. after 2014 manufactured ?
  - The data shows many vehicles data after 2014 manufactured
- #Can we find out data of only two wheelers from this data ? Which is the oldest bike sold here?
  - Yes, data of only two wheelers can be found, the oldest bike is Hero Super Splendor
- #Which is the newest bike sold here?
  - The data shows many new bike sold
- #Which is the most sold bike here?
  - The data shows many most sold bikes

YOUR SOLUTION AND ITS VALUE PROPOSITION

- #Do you find any deal in two wheelers which exceeded the general expectation ? Can you find reason for it ?
  - The answer to this question can be observed from a visualization to observe the data more accurately
- #Can we find out data of only cars from this data ?
  - Yes, the data of only cars can be found from this data
- #Which is the oldest car sold here?
  - The oldest car sold is wagon r
- #Which is the newest car sold here?
  - The newest car sold is land cruiser
- #Do you find any deal in cars which exceeded the general expectation ? Can you find reason for it ?
  - The answer to this question can be observed from a visualization to observe the data more accurately

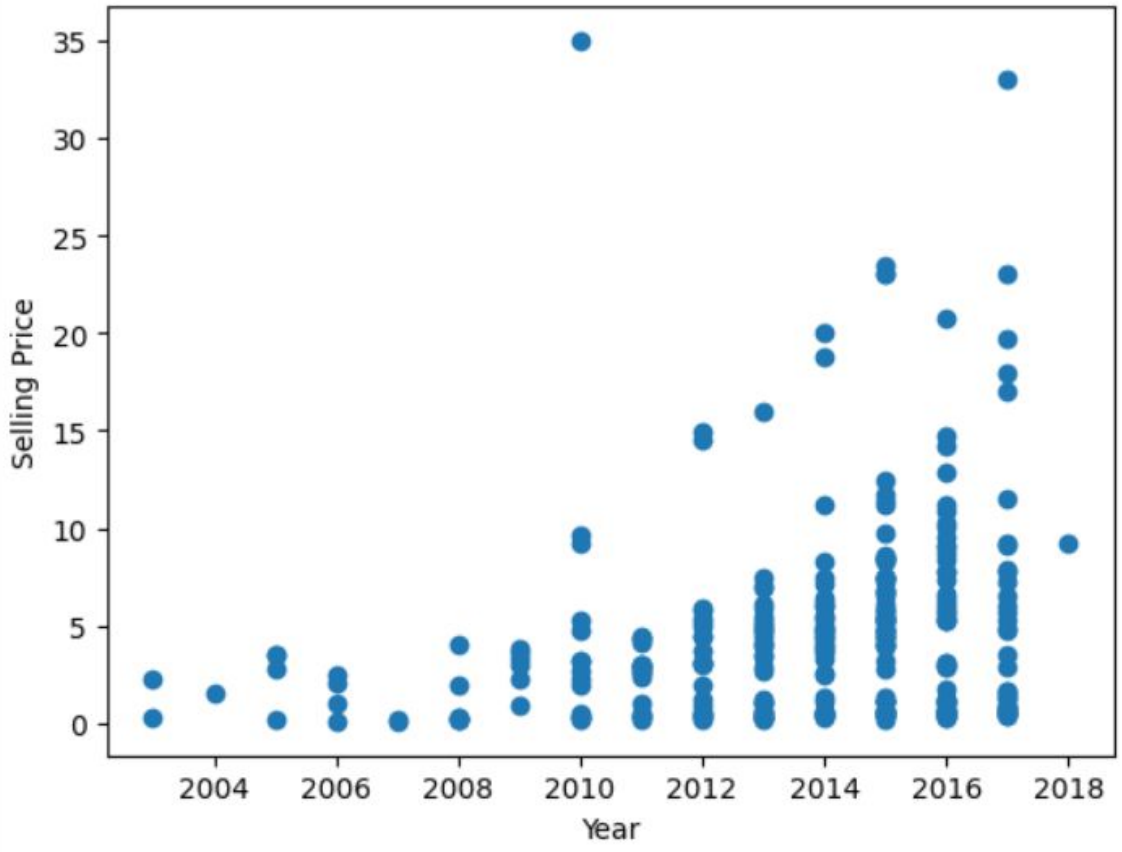
HOW DID YOU CUSTOMIZE THE PROJECT AND MAKE IT YOUR OWN

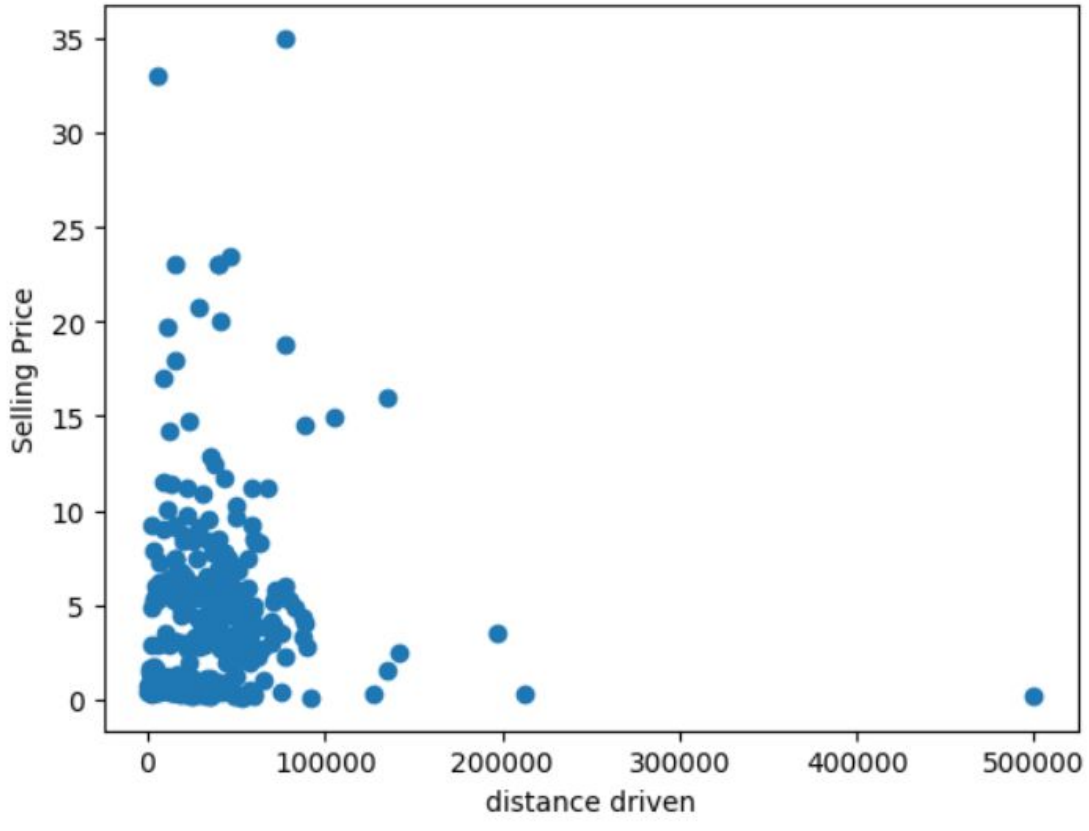
```
#In general selling price is affected by age of vehicle and distance driven by vehicle , is it observable from data ?
import matplotlib.pyplot as plt
plt.scatter(df['Year'], df['Selling_Price'])
plt.xlabel('Year')
plt.ylabel('Selling Price')
plt.show()
plt.scatter(df['Kms_Driven'], df['Selling_Price'])
plt.xlabel('distance driven')
plt.ylabel('Selling Price')
plt.show()
```

I made two scatter plots to show:

How selling price is affected by age

How selling price is affected by distance driven by vehicle





MODELLING

1. Loaded the dataset (CarDekho) in Google Collab
2. Cleaned the data and checked for missing values, null values
3. Imported libraries: Pandas, NumPy, Matplotlib to analyze the data
4. Analyzed the results to answer various questions to identify certain trends in the data
5. Created visualizations (box plots, scatter plots) to better understand the results from the data

RESULTS

```
#Which is the most and least cost depreciated vehicle in data ?
df['Depreciation'] = df['Present_Price'] - df['Selling_Price']

most_depreciated = df.loc[df['Depreciation'].idxmax(), 'Car_Name']
least_depreciated = df.loc[df['Depreciation'].idxmin(), 'Car_Name']

print(f"{most_depreciated} is the most cost depreciated vehicle")
print(f"{least_depreciated} is the least cost depreciated vehicle")
```

```
land cruiser is the most cost depreciated vehicle
Honda Activa 4G is the least cost depreciated vehicle
```

```
#Which is the most sold vehicle in this data ?
most_sold_vehicle = df['Car_Name'].value_counts().idxmax()
print(f"{most_sold_vehicle} is the most sold vehicle")
```

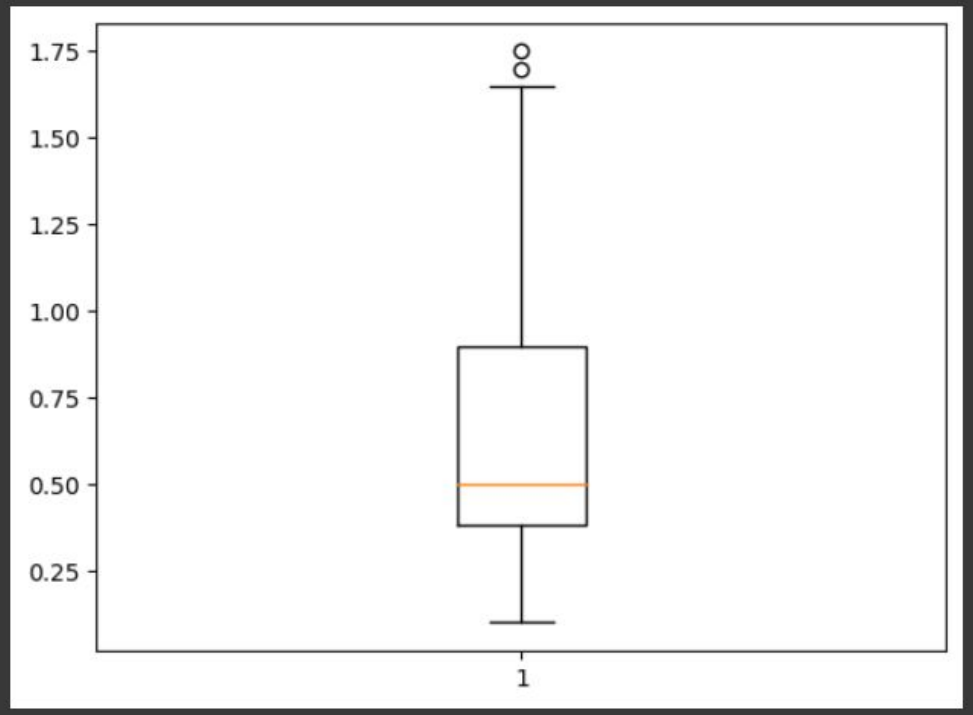
```
city is the most sold vehicle
```

```
#Can we find out data of only two wheelers from this data ? Which is the oldest bike sold here?
two_wheelers = df[df["Present_Price"]<3.5]
two_wheelers = two_wheelers.loc[two_wheelers.Car_Name != "alto k10"]
two_wheelers = two_wheelers.loc[two_wheelers.Car_Name != "800"]
two_wheelers = two_wheelers.loc[two_wheelers.Car_Name != "omni"]
two_wheelers.reset_index(drop=True,inplace=True)
two_wheelers.head()
two_wheelers.loc[two_wheelers.Year == two_wheelers.Year.min()]
```

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
89	Hero Super Splendor	2005	0.2	0.57	55000	Petrol	Individual	Manual	0

RESULTS

```
#Do you find any deal in two wheelers which exceeded the general expectation ? Can you find reason for it ?
import matplotlib.pyplot as plt
plt.boxplot(two_wheelers.Selling_Price)
plt.show()
two_wheelers[two_wheelers.Selling_Price>1.6]
```



	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	Royal Enfield Thunder 500	2016	1.75	1.90	3000	Petrol	Individual	Manual	0
1	UM Renegade Mojave	2017	1.70	1.82	1400	Petrol	Individual	Manual	0
2	KTM RC200	2017	1.65	1.78	4000	Petrol	Individual	Manual	0

```
#How many single person owned vehicles are there in this database ?
single_owner = df[df['Owner'] == 0].shape[0]
print(f"{single_owner} single person owned vehicles")
```

```
290 single person owned vehicles
```

LINKS

- https://drive.google.com/drive/folders/1OaC9eLP_RI2YrotEnVN1RBxFK6XTF5UX?usp=sharing