AI-DRIVEN EXPLORATION AND PREDICTION OF COMPANY REGISTRATION TRENDS WITH REGISTER OF COMPANIES(RoC)

PHASE4

PROJECT

PREPARED BY

A.PRIYANKA,

510521205029,

BHARATHIDASAN ENGINEERING COLLEGE,

PHASE4 PROJECT SUBMISSION.

INTRODUCTION

AI-driven exploration and prediction performing exploratory data analysis refers to the use of artificial intelligence (AI) techniques to automate and enhance the process of exploring and analyzing data to make predictions. It involves leveraging machine learning algorithms, data mining, and statistical methods to uncover hidden patterns, trends, and insights within a dataset, thereby aiding in the prediction of future outcomes or behavior. This approach combines the power of AI to assist in data preprocessing, feature selection, and modeling, making the exploratory data analysis more efficient and accurate.

DATASET

	CORPORAT	COMPANY_	COMPANY_	COMPANY_	COMPANY_	COMPANY_	DATE_OF_F	REGISTERED	AUTHORIZE	PAIDUP_CA	INDUSTRIAI	PRINCIPAL_	REGISTERED	REGISTRAR_	EMAIL_ADD	LATEST_YEA	LATEST
2	F00643	HOCHTIEFF	NAEF	NA	NA	NA	1/12/1961	Tamil Nadu	0	0	NA	Agriculture	AMBLE SIDE	ROC襟ELHI	NA	NA	NA
3	F00721	SUMITOMC	ACTV	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	FLAT NO. 6,	ROC襟ELHI	shuchi.chu	NA	NA
4	F00892	SRILANKAN	ACTV	NA	NA	NA	1/3/1982	Tamil Nadu	0	0	NA	Agriculture	SRILANKAN	ROC襟ELHI	shree16us@	NA	NA
5	F01208	CALTEX IND	NAEF	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	GOLD CRES	ROC襟ELHI	NA	NA	NA
6	F01218	GE HEALTH(	ACTV	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	FF-3 Palani	ROC襟ELHI	karthick999	NA	NA
7	F01265	CAIRN ENEI	NAEF	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	WELLINGT(	ROC襟ELHI	neerja.shar	NA	NA
8	F01269	TORIELLI S.I	ACTV	NA	NA	NA	5/9/1995	Tamil Nadu	0	0	NA	Agriculture	6, Mangaya	ROC襟ELHI	chennai@t	NA	NA
9	F01311	HARDY EXP	ACTV	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	5TH FLOOR,	ROC襟ELHI	venkatesh.	NA	NA
10	F01314	HOCHTIOF /	ACTV	NA	NA	NA	11/4/1996	Tamil Nadu	0	0	NA	Agriculture	NEW NO.8(	ROC襟ELHI	kumar@int	NA	NA
11	F01412	EPSON SIN(	ACTV	NA	NA	NA	25-04-1997	Tamil Nadu	0	0	NA	Agriculture	7C CEATUR'	ROC襟ELHI	NA	NA	NA
12	F01426	CARGOLUX	ACTV	NA	NA	NA	11/6/1997	Tamil Nadu	0	0	NA	Agriculture	OFFICE NO	ROC襟ELHI	NA	NA	NA
13	F01468	CHO HEUN(	NAEF	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	129, MANPI	ROC襟ELHI	chowelacc	NA	NA
14	F01543	NYCOMED /	ACTV	NA	NA	NA	27-10-1998	Tamil Nadu	0	0	NA	Agriculture	A D 46 1ST	ROC襟ELHI	NA	NA	NA
15	F01544	CHERRINGT	ACTV	NA	NA	NA	1/5/2000	Tamil Nadu	0	0	NA	Agriculture	10HADDOW	ROC襟ELHI	NA	NA	NA
16	F01563	SHIMADZU	NAEF	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	FIRST FLOO	ROC襟ELHI	kousik@vsr	NA	NA
17	F01565	CORK INTEI	ACTV	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	ARJAY APE)	ROC襟ELHI	NA	NA	NA
18	F01566	ERBIS ENG(	ACTV	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	39,2nd Mai	ROC襟ELHI	NA	NA	NA
19	F01589	RALF SCHNI	NAEF	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	FLAT C, 'SAI	ROC襟ELHI	NA	NA	NA
20	F01593	MITRAJAYA	ACTV	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	OLD NO 14	ROC襟ELHI	NA	NA	NA
21	F01618	HEAT AND (	ACTV	NA	NA	NA	13-07-1999	Tamil Nadu	0	0	NA	Agriculture	A40 OLD N(	ROC襟ELHI	ncrajagopal	NA	NA
22	F01628	DIREX SYST	ACTV	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	F-1, FIRST F	ROC襟ELHI	direx@vsnl	NA	NA
23	F01641	NMB-MINE	NAEF	NA	NA	NA	NA	Tamil Nadu	0	0	NA	Agriculture	Level - 2 Re	ROC襟ELHI	stsogawa@	NA	NA

OVERVIEW

Personality test

EXPLORATORY DATA ANALYSIS:

AI-driven exploration and prediction performing exploratory data analysis refers to the use of artificial intelligence (AI) techniques to automate and enhance the process of exploring and analyzing data to make predictions. It involves leveraging machine learning algorithms, data mining, and statistical methods to uncover hidden patterns, trends, and insights within a dataset, thereby aiding in the prediction of future outcomes or behavior. This approach combines the power of AI to assist in data preprocessing, feature selection, and modeling, making the exploratory data analysis more efficient and accurate.

FEATURE ENGINEERING:

AI-driven exploration and prediction performing feature engineering involves using artificial intelligence techniques to automatically create, select, or transform features (variables) in a dataset to improve the performance of machine learning models. Feature engineering is a critical step in the machine learning pipeline, and AI-driven methods can help identify relevant features, reduce dimensionality, and generate new attributes from the existing data. This process aims to enhance the predictive capabilities of AI models by optimizing the input data, ultimately leading to better model accuracy and effectiveness in solving specific tasks.

PREDICTIVE MODELING:

AI-driven exploration and prediction performing predictive modeling refers to the application of artificial intelligence and machine learning techniques to build and deploy models that make predictions based on historical or existing data. In this process, AI algorithms analyze the data to identify patterns, relationships, and trends, and use this information to forecast future outcomes or make informed decisions. The predictive models created through this approach can be used in various domains, such as finance, healthcare, marketing, and more, to make predictions about customer behavior, stock prices, disease diagnoses, or any other relevant forecasting task. AI-driven predictive modeling automates and enhances the accuracy of prediction tasks by leveraging advanced machine learning algorithms and data-driven insights.

STEPS FOR EXPLORATORY DATA ANALYSIS

1. Data Collection
2. Data Preprocessing
3. Feature Selection
4. Data Visualization
5. Descriptive Statistics
6. Correlation Analysis
7. Outlier Detection
8. Data Exploration
9. Hypothesis Testing
10. Feature Engineering
11. Model Training
12. Model Evaluation
13. Hyperparameter Tuning
14. Model Interpretability
15. Model Deployment
16. Continuous Monitoring

STEPS FOR FEATURE ENGINEERING

1. Data Collection
2. Data Preprocessing
3. Feature Selection
4. Data Visualization
5. Descriptive Statistics
6. Correlation Analysis
7. Outlier Detection
8. Data Exploration
9. Hypothesis Testing
10. Feature Engineering
11. Model Training
12. Model Evaluation
13. Hyperparameter Tuning
14. Model Interpretability
15. Model Deployment
16. Continuous Monitoring

STEPS FOR PREDICTIVE MODELING

1.Data Collection
2.Data Preprocessing
3.Feature Engineering
4.Data Splitting
5.Model Selection
6.Model Training
7.Hyperparameter Tuning
8Model Evaluation
9.Model Interpretability
10.Cross-Validation
11.Model Fine-Tuning
12.Ensemble Methods
13.Model Validation
14.Model Deployment
15.Continuous Monitoring
16.Feedback Loop

STEPS FOR PREDICTIVE MODELING

1.Data Collection
2.Data Preprocessing
3.Feature Engineering
4.Data Splitting
5.Model Selection
6.Model Training
7.Hyperparameter Tuning
8Model Evaluation
9.Model Interpretability
10.Cross-Validation
11.Model Fine-Tuning
12.Ensemble Methods
13.Model Validation
14.Model Deployment
15.Continuous Monitoring
16.Feedback Loop

PROGRAM

Pandas and Numpy have been used for Data Manipulation and numerical Calculations

Pandas and Numpy have been used for Data Manipulation and numerical Calculations
Matplotlib and Seaborn have been used for Data visualizations.

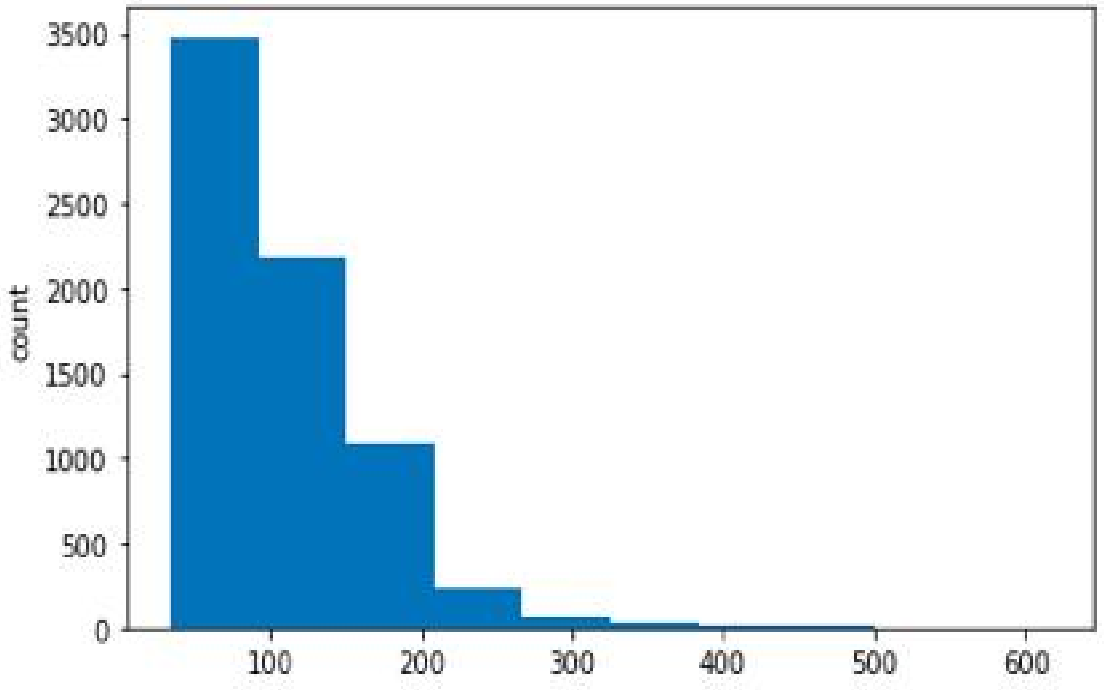
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#to ignore warnings
import warnings
warnings.filterwarnings('ignore')
data = pd.read_csv("used_cars.csv")
data.head()
```

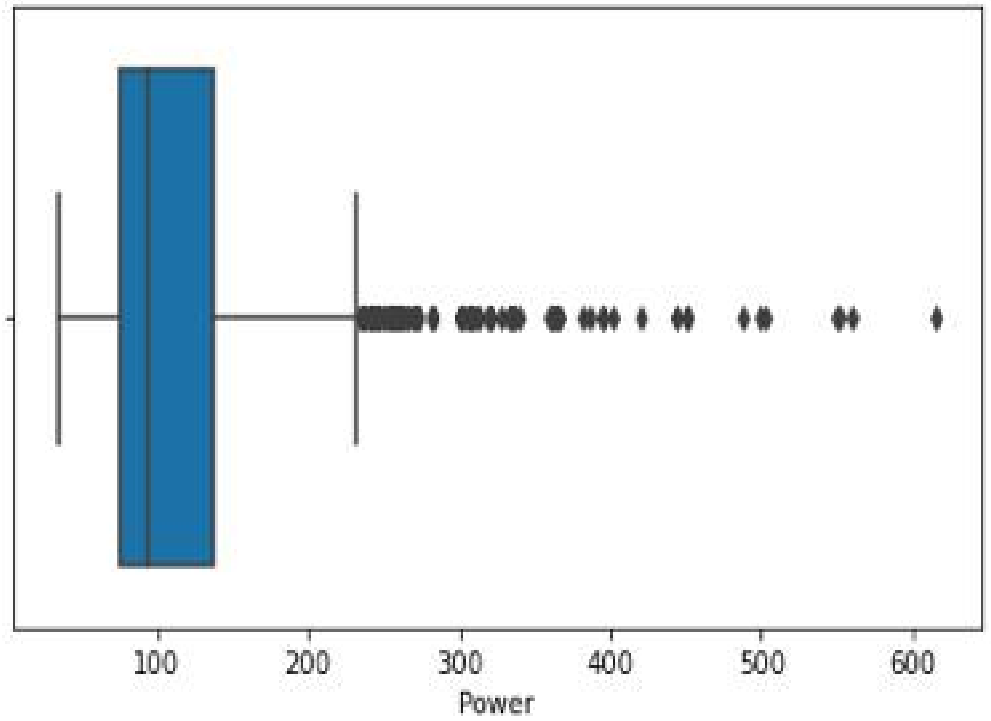
	S.No.	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_price	Price
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.60	998.0	58.16	5.0	NaN	1.75
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	NaN	12.50
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.20	1199.0	88.70	5.0	8.61	4.50
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	NaN	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	NaN	17.74

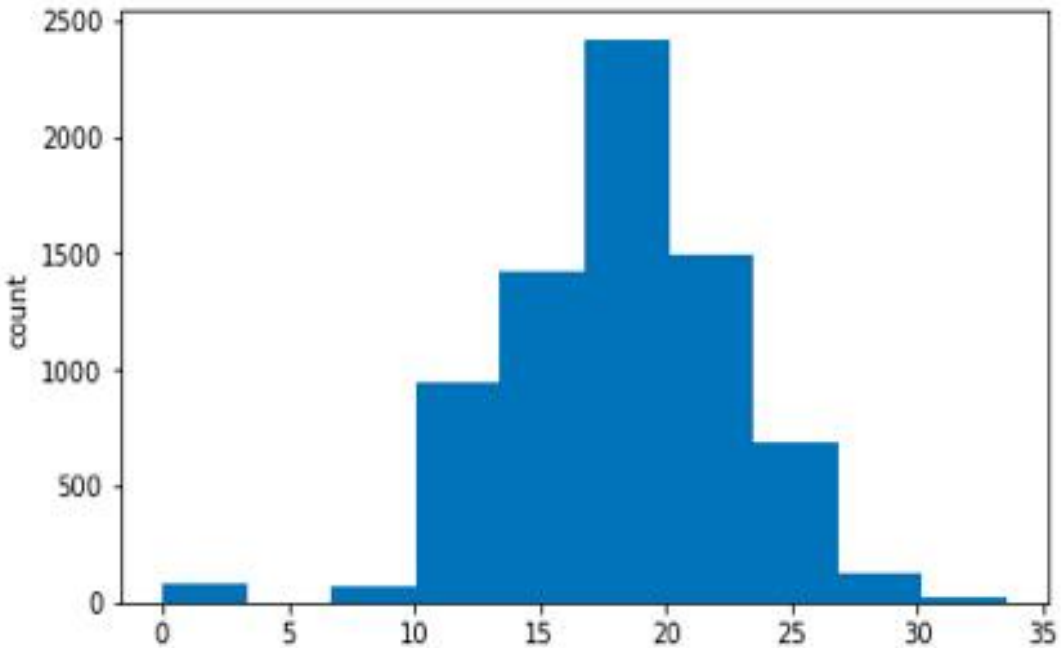
```
cat_cols=data.select_dtypes(include=['object']).columns
num_cols =
data.select_dtypes(include=np.number).columns.tolist()
print("Categorical Variables:")
print(cat_cols)
print("Numerical Variables:")
print(num_cols)
```

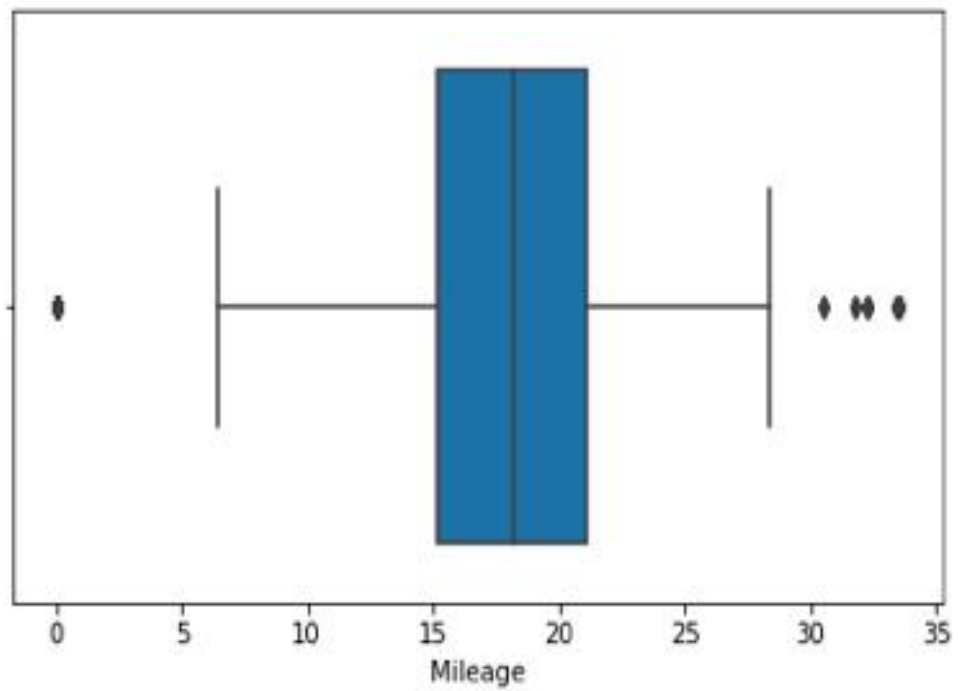
```
Categorical Variables:
Index(['Name', 'Location', 'Fuel_Type', 'Transmission', 'Owner_Type', 'Brand',
       'Model'],
      dtype='object')
Numerical Variables:
['Year', 'Kilometers_Driven', 'Mileage', 'Engine', 'Power', 'Seats', 'New_price', 'Price', 'Car_Age']
```

```
for col in num_cols:
print(col)
print('Skew :'
, round(data[col].skew(), 2))
plt.figure(figsize = (15, 4))
plt.subplot(1, 2, 1)
data[col].hist(grid=False)
plt.ylabel('count')
plt.subplot(1, 2, 2)
sns.boxplot(x=data[col])
plt.show()
```

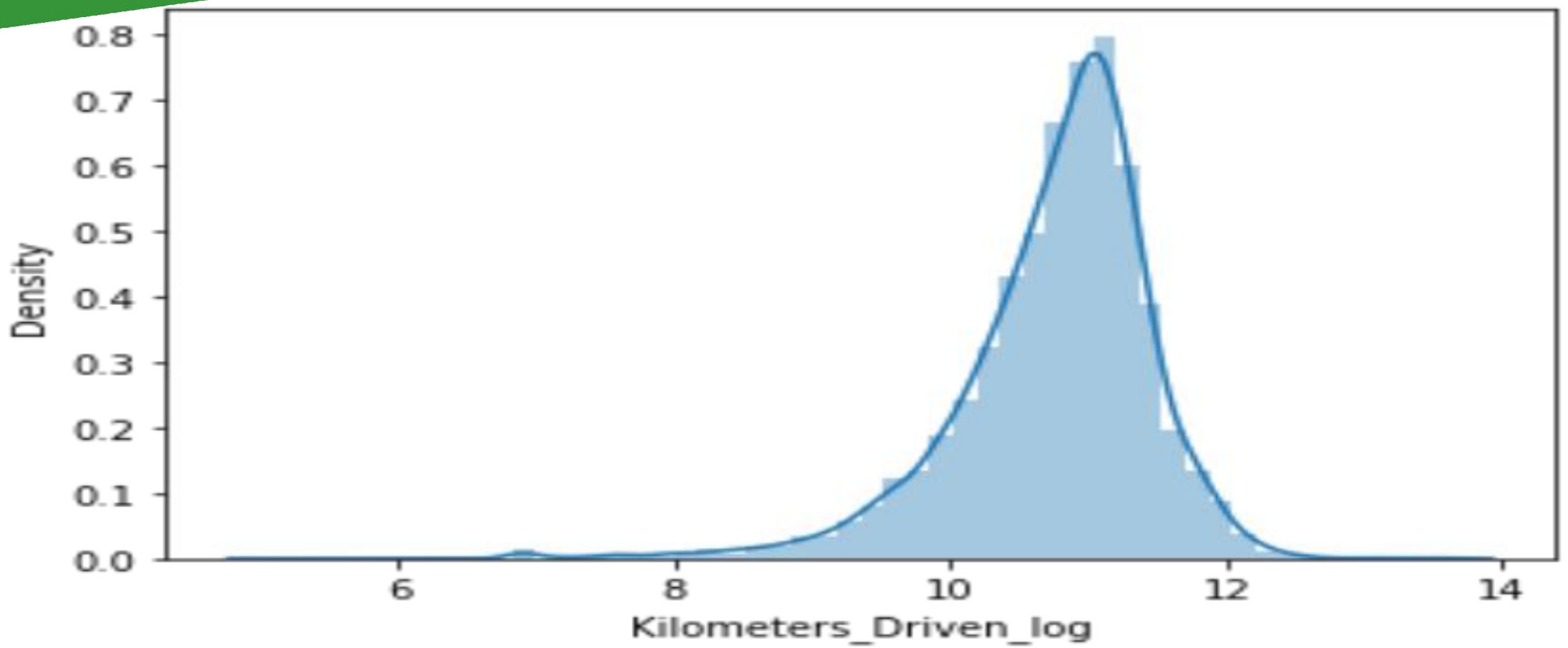

3500
3000
2500
2000
1500
1000
500
0
count
100
200
300
400
500
600


100
200
300
400
500
600
Power


2500
2000
1500
1000
500
0
count
0
5
10
15
20
25
30
35

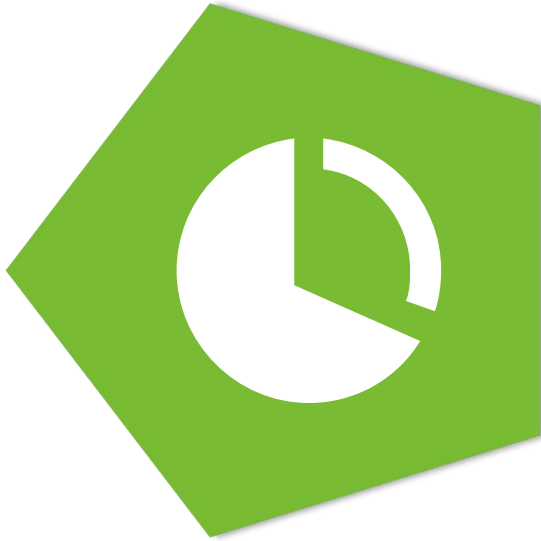

0
5
10
15
20
25
30
35
Mileage

```
# Function for log transformation of the column
def log_transform(data,col):
for colname in col:
if (data[colname] == 1.0).all():
data[colname + '_log'] = np.log(data[colname]+1)
else:
data[colname + '_log'] = np.log(data[colname])
data.info()
log_transform(data,['Kilometers_Driven'
,
'Price'])
#Log transformation of the feature 'Kilometers_Driven'
sns.distplot(data["Kilometers_Driven_log"], axlabel=
"Kilometers_Driven_log"
```


Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
6
8
10
12
14
Kilometers_Driven_log

CONCLUSION

In conclusion, AI-driven exploration and prediction represent a transformative approach to data analysis and forecasting. By harnessing the power of artificial intelligence, this methodology enhances the efficiency, accuracy, and depth of insights derived from data. Through advanced techniques in data preprocessing, feature engineering, and predictive modeling, AI enables us to uncover hidden patterns, make informed predictions, and drive data-driven decision-making in a wide range of applications.



This approach has the potential to revolutionize industries such as finance, healthcare, marketing, and many others, by automating and optimizing the data analysis process. By continuously learning from data and adapting to changing circumstances, AI-driven exploration and prediction systems provide a valuable tool for organizations seeking to gain a competitive edge, improve efficiency, and make more informed choices.

New York University

Thank you so much for your watching

My name：Aaron

My major：