Project Report

On

Car Sales Prediction

Submitted in partial fulfilment of the requirements for the award of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

(Artificial Intelligence & Machine Learning)

by

Ms.YASHASWINY SRIPADA-22WH1A6607

Ms.R.ISHWARYA – 22WH1A6609

Ms.S.AISHWARYA– 22WH1A6644

Ms.N.VAISHNAVI– 22WH1A6645

Under the esteemed guidance of

Ms. A Naga Kalyani

Assistant Professor, CSE(AI&ML)

BVRIT HYDERABAD College of Engineering for Women

(UGC Autonomous Institution | Approved by AICTE | Affiliated to JNTUH)

(NAAC Accredited - A Grade | NBA Accredited B.Tech. (EEE, ECE, CSE and IT)

Bachupally, Hyderabad – 500090

2024-25

Department of Computer Science & Engineering

(Artificial Intelligence & Machine Learning)

BVRIT HYDERABAD COLLEGE OF ENGINEERING FOR WOMEN

(Approved by AICTE, New Delhi and Affiliated to JNTUH, Hyderabad)

Accredited by NBA and NAAC with A Grade

Bachupally, Hyderabad – 500090

2024-25

CERTIFICATE

This is to certify that the major project entitled **"Car Sales prediction using python"** is a bonafide work carried out **by Ms. Yashaswiny Sripada (22wh1a6607), Ms. R.Ishwarya (22wh1a6609), Ms. S.Aishwarya (22wh1a6644), Ms. N.Vaishnavi (22wh1a6645)** in partial fulfillment for the award of B. Tech degree in **Computer Science & Engineering (AI&ML), BVRIT HYDERABAD College of Engineering for Women, Bachupally, Hyderabad,** affiliated to Jawaharlal Nehru Technological University Hyderabad, Hyderabad under my guidance and supervision. The results embodied in the project work have not been submitted to any other University or Institute for the award of any degree or diploma.

Supervisor
Ms. A Naga Kalyani
Assistant Professor

Dept of CSE(AI&ML)

Head of the Department
Dr. B. Lakshmi Praveena
HOD & Professor

Dept of CSE(AI&ML)

External Examiner

DECLARATION

We hereby declare that the work presented in this project entitled **"Car Sales prediction using python"** submitted towards completion of Project work in IV Year of B.Tech of CSE(AI&ML) at **BVRIT HYDERABAD College of Engineering for Women,** Hyderabad is an authentic record of our original work carried out under the guidance of **Ms. A Naga Kalyani, Assistant Professor, Department of CSE(AI&ML).**

Sign with Date:

Yashaswiny Sripada

(22wh1a6607)

Sign with Date:

R.Ishwarya

(22wh1a6607)

Sign with Date:

S.Aishwarya

(22wh1a6644)

Sign with Date:

N.Vaishnavi

(22wh1a6645)

ACKNOWLEDGEMENT

We would like to express our sincere thanks to **Dr. K. V. N. Sunitha, Principal, BVRIT HYDERABAD College of Engineering for Women**, for her support by providing the working facilities in the college.

Our sincere thanks and gratitude to **Dr. B. Lakshmi Praveena, Head of the Department, Department of CSE(AI&ML), BVRIT HYDERABAD College of Engineering for Women,** for all timely support and valuable suggestions during the period of our project.

We are extremely thankful to our Internal Guide**, Ms. A Naga Kalyani, Assistant Professor, CSE(AI&ML), BVRIT HYDERABAD College of Engineering for Women,** for her constant guidance and encouragement throughout the project.

Finally, we would like to thank our Major Project Coordinator, all Faculty and Staff of CSE(AI&ML) department who helped us directly or indirectly. Last but not least, we wish to acknowledge our **Parents and Friends** for giving moral strength and constant encouragement.

Yashaswiny Sripada(22wh1a6607)

R.Ishwarya (22wh1a6609)

S.Aishwarya(22wh1a6644)

N.Vaishnavi(22wh1a6645)

ABSTRACT

This project analyzes car sales data using Python to provide insights into sales trends, popular car models, and pricing patterns. The report utilizes data manipulation libraries such as Pandas for organizing and processing data, along with Matplotlib and Seaborn for visualization. Key metrics like total sales, average prices, and sales distribution by brand or model are calculated. The project also explores relationships between variables like car price and features, enabling data-driven insights. The analysis aims to assist stakeholders in understanding market trends, optimizing inventory, making informed decisions to improve sales performance and customer satisfaction.

PROBLEM STATEMENT

The automotive industry generates vast amounts of sales data, which often remain underutilized due to the lack of efficient analytical tools. Businesses face challenges in identifying trends, understanding customer preferences, and optimizing inventory and marketing strategies. Key issues include:

1.Unorganized Data: Difficulty in extracting meaningful insights from raw and complex datasets.
2.Lack of Visualization: Absence of clear and intuitive representations of sales trends, customer behaviour, and regional performance.
3.Inefficient Decision-Making: Limited ability to identify top-selling models, peak sales periods, and market demand due to inadequate analysis.

This project aims to address these challenges by developing a Car Sale Report using Python, which will provide a structured, visual, and actionable understanding of car sales data. The solution will empower stakeholders to make data-driven decisions and improve overall business efficiency.

DATA SET

Car Sales prediction – Kaggle

https://www.kaggle.com/datasets/missionjee/car-sales-report

SOURCE CODE

```
1#import pandas as pd
file_path = '/content/Car Sales.xlsx - car_data.csv'
data = pd.read_csv(file_path)

2# print(data.head())

 3#   from sklearn.model_selection import train_test_split
  from sklearn.ensemble import RandomForestRegressor
  from sklearn.metrics import mean_squared_error
  import pandas as pd

  # Data preprocessing
  data['Engine'] = data['Engine'].astype('category').cat.codes
  data['Transmission'] = data['Transmission'].astype('category').cat.codes
  data['Company'] = data['Company'].astype('category').cat.codes
  data['Body Style'] = data['Body Style'].astype('category').cat.codes
  data['Color'] = data['Color'].astype('category').cat.codes

  X = data[['Engine', 'Transmission', 'Company', 'Body Style', 'Color']]
  y = data['Price ($)']

  # Split the data
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
  random_state=42)

  # Train the model
  model = RandomForestRegressor()
  model.fit(X_train, y_train)
```

```
# Evaluate the model
predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print(f"Mean Squared Error: {mse}")
```

```
4# from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report

# Create a binary target variable for price
threshold = 25000
data['Price_Label'] = (data['Price ($)'] > threshold).astype(int)

# Encode categorical features
data['Engine'] = data['Engine'].astype('category').cat.codes
data['Transmission'] = data['Transmission'].astype('category').cat.codes
data['Company'] = data['Company'].astype('category').cat.codes
data['Body Style'] = data['Body Style'].astype('category').cat.codes

# Select features and target
X = data[['Annual Income', 'Engine', 'Transmission', 'Company', 'Body Style']]
y = data['Price_Label']

# Split the data into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train a logistic regression model
model = LogisticRegression(max_iter=1000)
```

```
model.fit(X_train, y_train)

# Evaluate the model
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
print("Classification Report:")
print(classification_report(y_test, y_pred))
5#from sklearn.neighbors import NearestNeighbors
import numpy as np

# Features for recommendation
X = data[['Annual Income', 'Price ($)']]

# Fit Nearest Neighbors model
model = NearestNeighbors(n_neighbors=5, metric='euclidean')
model.fit(X)

# Example query for recommendation
query = np.array([[50000, 20000]])
distances, indices = model.kneighbors(query)

# Recommended cars
recommendations = data.iloc[indices[0]]
print(recommendations[['Company', 'Model', 'Price ($)']])
6# [1:01 pm, 5/12/2024] Ishwarya Ravulapenta: from sklearn.decomposition
import PCA
from sklearn.preprocessing import StandardScaler
```

```
X = data[['Annual Income', 'Price ($)']]

# Standardize the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Apply PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

print("Explained Variance Ratio:", pca.explained_variance_ratio_)
[1:02 pm, 5/12/2024] Ishwarya Ravulapenta: from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
X = data[['Annual Income']]
# Apply KMeans clustering
kmeans = KMeans(n_clusters=3, random_state=42)
data['Cluster'] = kmeans.fit_predict(X)
# Visualize the clusters
plt.scatter(X['Annual Income'], [0]*len(X), c=data['Cluster'], cmap='viridis')
plt.title('Customer Segmentation')
plt.xlabel('Annual Income')
plt.show()
7# import matplotlib.pyplot as plt
data['Date'] = pd.to_datetime(data['Date'])
data.set_index('Date', inplace=True)
# Aggregate monthly sales
monthly_sales = data['Price ($)'].resample('M').sum()

# Plot sales trends
```

```
plt.figure(figsize=(10, 6))
monthly_sales.plot()
plt.title('Monthly Sales Trends')
plt.xlabel('Date')
plt.ylabel('Total Sales ($)')
plt.show()
```

8# [1:02 pm, 5/12/2024] Ishwarya Ravulapenta: import matplotlib.pyplot as plt

```
# Calculating total sales for each company
company_sales = data.groupby('Company')['Price ($)'].sum().sort_values(ascending=False)

# Plot the bar chart
plt.figure(figsize=(10, 6))
company_sales.plot(kind='bar', color='skyblue')
plt.title('Total Sales by Car Company')
plt.xlabel('Car Company')
plt.ylabel('Total Sales ($)')
plt.xticks(rotation=45)
plt.show()
```

[1:02 pm, 5/12/2024] Ishwarya Ravulapenta: # Counting the number of cars sold by body style

```
body_style_counts = data['Body Style'].value_counts()
# Plot the pie chart
plt.figure(figsize=(8, 8))
body_style_counts.plot(kind='pie', autopct='%1.1f%%', startangle=140, colors=['lightblue', 'orange', 'green', 'red', 'purple'])
plt.title('Car Sales by Body Style')
plt.ylabel('')
plt.show()
```

OUTPUT

	Car_id	Date	Customer Name	Gender	Annual Income
0	C_CND_000001	1/2/2022	Geraldine	Male	13500
1	C_CND_000002	1/2/2022	Gia	Male	1480000
2	C_CND_000003	1/2/2022	Gianna	Male	1035000
3	C_CND_000004	1/2/2022	Giselle	Male	13500
4	C_CND_000005	1/2/2022	Grace	Male	1465000

	Dealer_Name	Company	Model
0	Buddy Storbeck's Diesel Service Inc	Ford	Expedition
1	C & M Motors Inc	Dodge	Durango
	Capitol KIA	Cadillac	Eldorado
3	Chrysler of Tri-Cities	Toyota	Celica
4	Chrysler Plymouth	Acura	TL

	Engine	Transmission	Color	Price ($)	Dealer_No
0	DoubleÂ Overhead Camshaft	Auto	Black	26000	06457-3834
1	DoubleÂ Overhead Camshaft	Auto	Black	19000	60504-7114
2	Overhead Camshaft	Manual	Red	31500	38701-8047
3	Overhead Camshaft	Manual	Pale Whit	14000	99301-3882
4	DoubleÂ Overhead Camshaft	Auto	Red	24500	53546-9427

	Body Style	Phone	Dealer_Region
0	SUV	8264678	Middletown
1	SUV	6848189	Aurora
2	Passenger	7298798	Greenville
3	SUV	6257557	Pasco
4	Hatchback	7081483	Janesville

Mean Squared Error

Mean Squared Error: 94325232.77855274

logistic regression model

Accuracy: 0.5696361355081556

Classification Report:

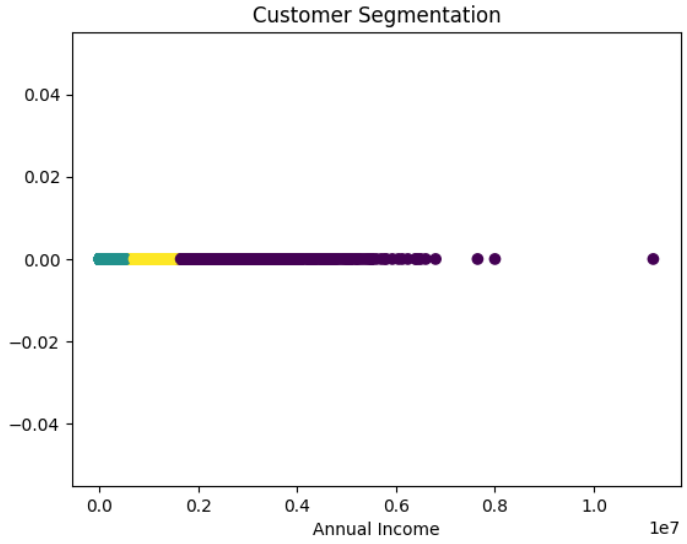
	precision	recall	f1-score	support
0	0.57	1.00	0.73	2724
1	0.00	0.00	0.00	2058
accuracy			0.57	4782
macro avg	0.28	0.50	0.36	4782
weighted avg	0.32	0.57	0.41	4782

	Company	Model	Price ($)
23068	7	Ram Pickup	20000
21344	21	Voyager	20000
22319	5	Cavalier	20000
20657	19	Frontier	20000
22251	12	S-Type	20000

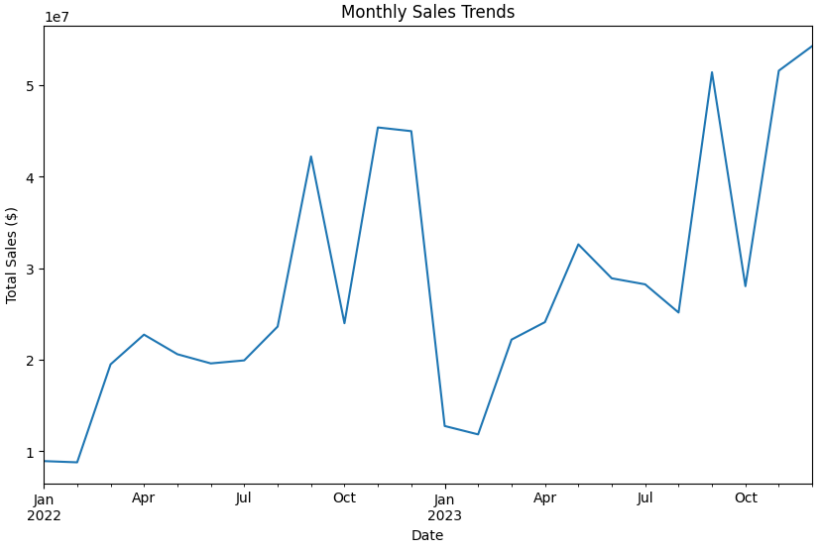
Explained Variance Ratio

Explained Variance Ratio: [0.50603248 0.49396752]

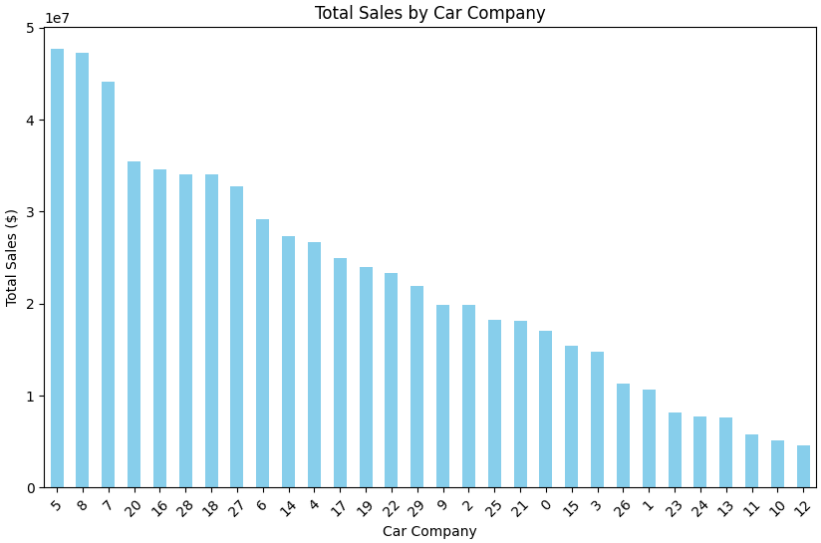
Customer Segmention



Monthly Sales Trends



Total Sales by Car Company



Car Sales by Body Style