



CAPSTONE PROJECT

COFFEE SALES PREDICTION USING MACHINE LEARNING

PRESENTED BY

STUDENT NAME: HARSHITA AGARWAL

COLLEGE NAME: IGNOU

DEPARTMENT: MANAGEMENT

EMAIL ID: harshitaaggarwal1995@gmail.com

AICTE STUDENT ID:



OUTLINE

- **PROBLEM STATEMENT**
- **PROPOSED SYSTEM/SOLUTION**
- **SYSTEM DEVELOPMENT APPROACH**
- **ALGORITHM & DEPLOYMENT**
- **SALES BY WEEKDAY**
- **HOURLY SALES TRENDS**
- **REVENUE BY COFFEE PRODUCT**
- **MACHINE LEARNING MODEL EVALUATION**
- **KEY FINDINGS & BUSINESS APPLICATIONS**
- **CONCLUSION**
- **REFERENCES**

PROBLEM STATEMENT

UNDERSTANDING COFFEE SALES TRENDS IS CRUCIAL FOR OPTIMIZING STOCK REFILLING, PRICING STRATEGIES, AND CUSTOMER SATISFACTION IN VENDING MACHINES. THIS PROJECT AIMS TO PREDICT COFFEE SALES USING **HISTORICAL VENDING MACHINE DATA** TO IMPROVE BUSINESS OPERATIONS.

PROPOSED SOLUTION

- **DATA COLLECTION:** TRANSACTION DETAILS FROM VENDING MACHINE SALES.
- **DATA PREPROCESSING:** HANDLING MISSING VALUES, FEATURE ENGINEERING (MONTH, DAY, HOUR).
- **MACHINE LEARNING MODEL:** LINEAR REGRESSION TO PREDICT COFFEE SALES.
- **DEPLOYMENT STRATEGY:** USING PREDICTIVE INSIGHTS FOR INVENTORY OPTIMIZATION.

SYSTEM APPROACH

- **LIBRARIES USED:** PANDAS, NUMPY, MATPLOTLIB, SEABORN, SKLEARN.
- **FEATURE ENGINEERING:** EXTRACTING MONTH, DAY, HOUR, CASH_TYPE, COFFEE_NAME.
- **MODEL TRAINING:** SPLITTING DATA (80% TRAINING, 20% TESTING).

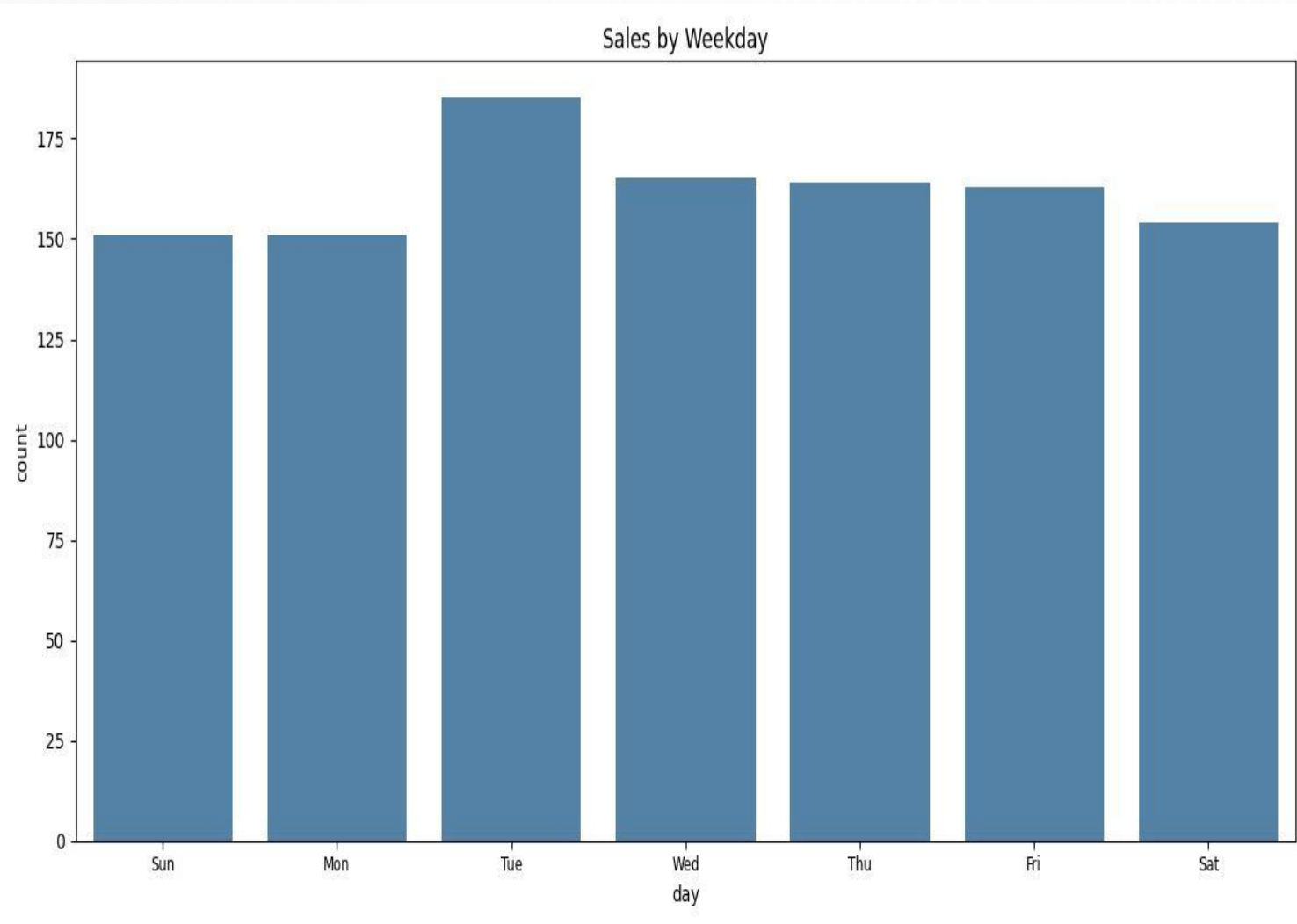
```
coffee_sales.py > ...
1  import numpy as np
2  import pandas as pd
3  import matplotlib.pyplot as plt
4  import seaborn as sns
5  import datetime as dt
6  import warnings
7  import os
8
9  from sklearn.model_selection import train_test_split
10 from sklearn.linear_model import LinearRegression
11 from sklearn.metrics import mean_squared_error, r2_score
12
13
14 warnings.filterwarnings('ignore')
15
16 # List input files (if applicable)
17 #for dirname, _, filenames in os.walk('/kaggle/input'):
18 |   #for filename in filenames:
19 |       # print(os.path.join(dirname, filename))
20
21 # Load data
22 coffee_data = pd.read_csv('E:\\AICETE & Edunet\\Coffee Sales project details\\coffee_sales.csv')
23
24 # Data cleaning
25 coffee_data['date'] = pd.to_datetime(coffee_data['date'])
26 coffee_data['datetime'] = pd.to_datetime(coffee_data['datetime'])
27 coffee_data['month'] = coffee_data['date'].dt.strftime('%Y-%m')
28 coffee_data['day'] = coffee_data['date'].dt.strftime('%w')
29 coffee_data['hour'] = coffee_data['datetime'].dt.strftime('%H')
30
```

ALGORITHM & DEPLOYMENT

- **ALGORITHM:** LINEAR REGRESSION TRAINED ON EXTRACTED FEATURES.
- **MODEL TRAINING:** USED HISTORICAL SALES PATTERNS TO LEARN DEMAND TRENDS.
- **PREDICTION PROCESS:** ESTIMATED FUTURE COFFEE SALES BASED ON EXTRACTED FEATURES.

```
coffee_sales.py > ...
111
112 # Prepare dataset for ML
113
114 # Select features (exclude 'datetime', 'date', 'card' for now)
115 features = ['month', 'day', 'hour', 'cash_type', 'coffee_name']
116
117 # One-hot encode categorical features
118 X = pd.get_dummies(coffee_data[features], drop_first=True)
119
120 # Target variable
121 y = coffee_data['money']
122
123 # Split data
124 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
125
126 # Initialize and train model
127 model = LinearRegression()
128 model.fit(X_train, y_train)
129
130 # Predict
131 y_pred = model.predict(X_test)
132
133 # Evaluation
134 mse = mean_squared_error(y_test, y_pred)
135 r2 = r2_score(y_test, y_pred)
136
137 print(f"Model Evaluation:\nMean Squared Error: {mse:.2f}\nR2 Score: {r2:.2f}")
138
139 # Coefficients for interpretation
140 coefficients = pd.DataFrame({'Feature': X.columns, 'Coefficient': model.coef_})
141 print("\nModel Coefficients:")
142 print(coefficients.sort_values(by='Coefficient', ascending=False))
143
144 # Save model coefficients
145 coefficients.to_csv('model_coefficients.csv', index=False)
```


SALES BY WEEKDAY

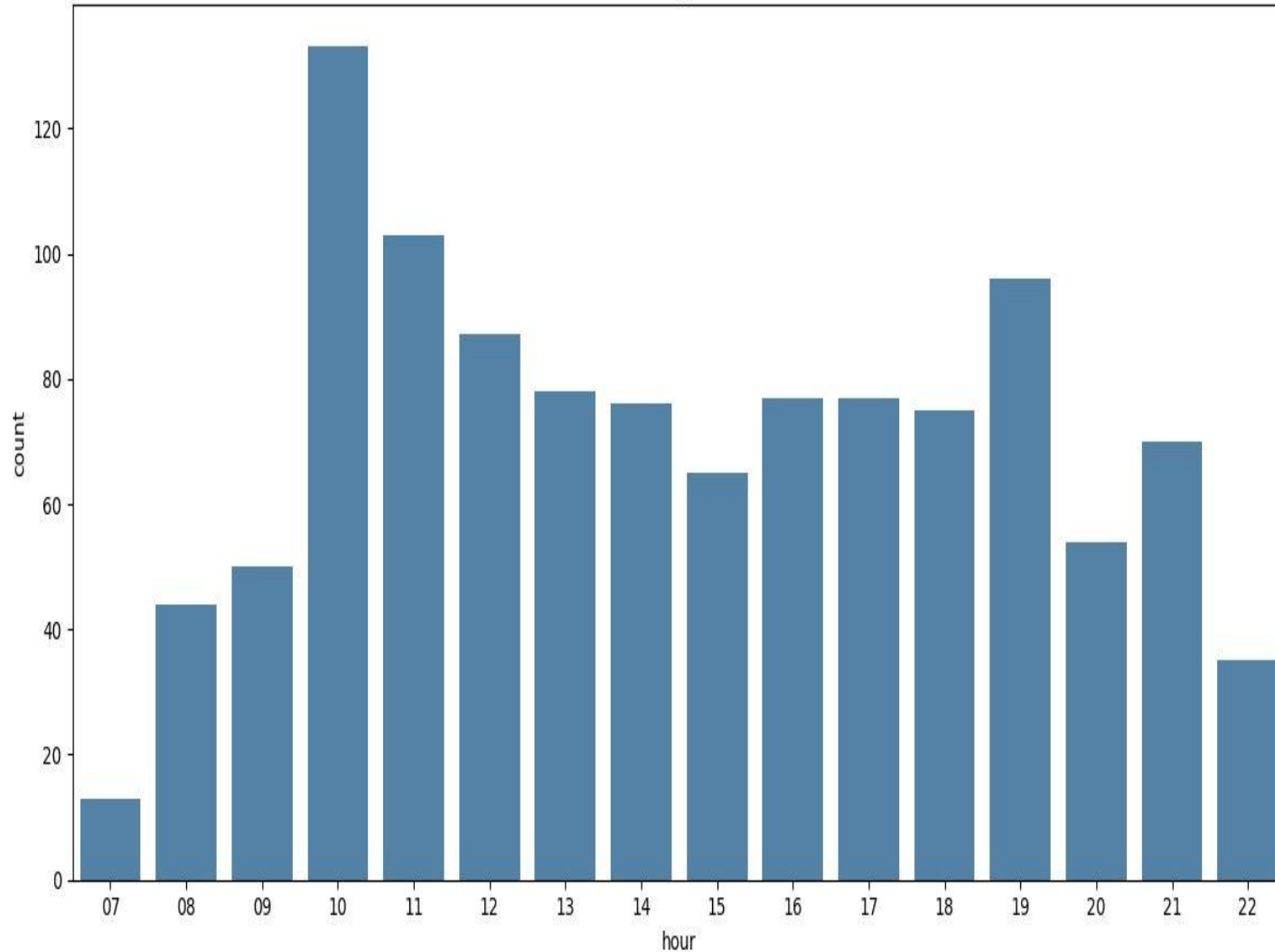


KEY INSIGHTS:

- **TUESDAY:** HIGHEST SALES.
- **MONDAY:** LOWEST DEMAND.
- **BUSINESS IMPACT:** ADJUST RESTOCKING SCHEDULE BASED ON WEEKDAY TRENDS.

HOURLY SALES TRENDS

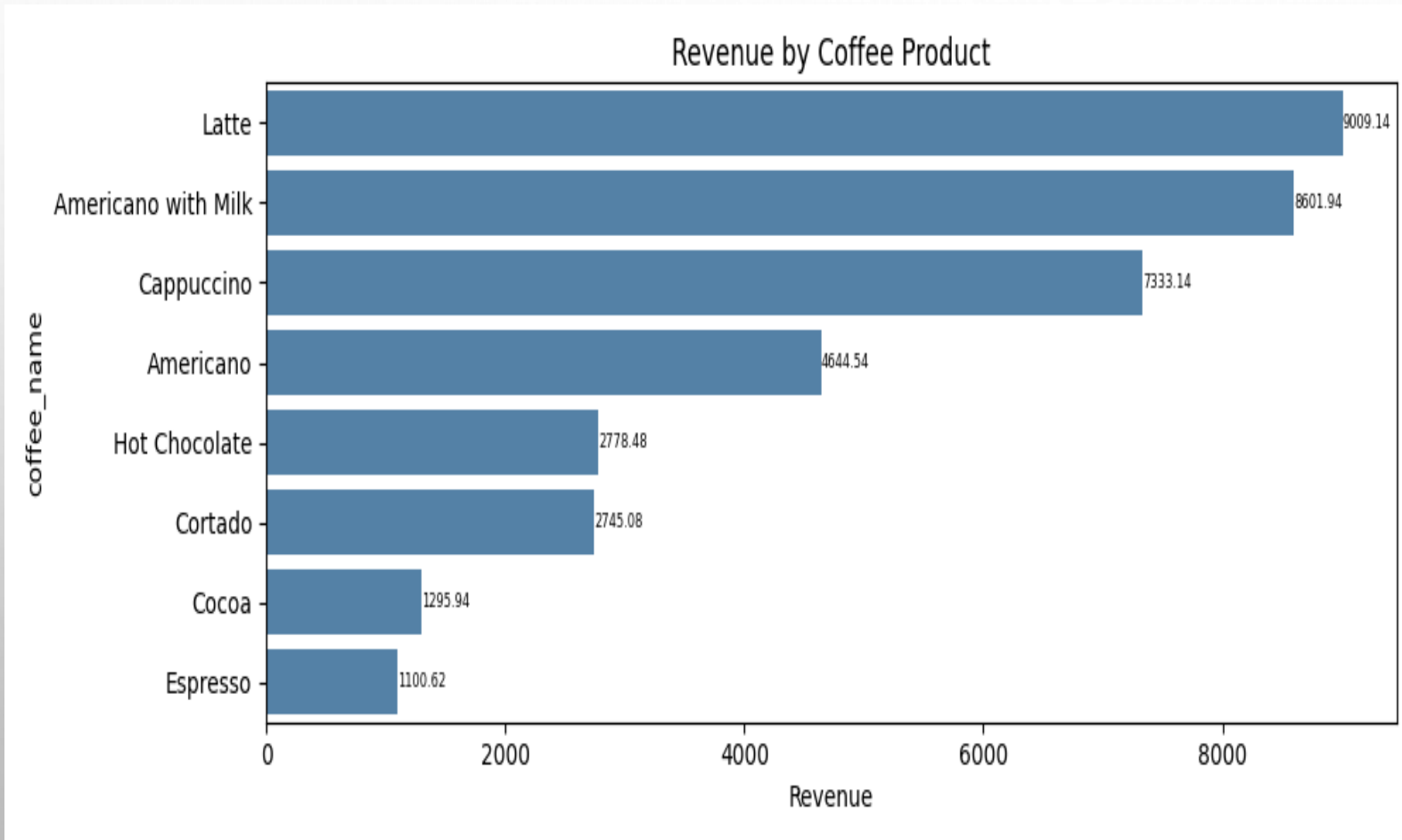
Sales by Hour



KEY FINDINGS:

- **PEAK HOUR:** 10:00 AM (HIGHEST DEMAND).
- **EVENING PEAKS:** 16:00, 17:00, 19:00 PM.
- **BUSINESS IMPACT:** OPTIMIZE STOCK AVAILABILITY DURING PEAK SALES TIMES.

REVENUE BY COFFEE PRODUCT



- **TOP SELLING COFFEES:**
 - LATTE (₹9009.14), AMERICANO WITH MILK (₹8601.94), CAPPUCCINO (₹7333.14).
- **LEAST SELLING:** COCOA, ESPRESSO.
- **BUSINESS IMPACT:** OPTIMIZE INVENTORY BASED ON HIGHEST REVENUE.

MACHINE LEARNING MODEL EVALUATION

```
PS E:\AICETE & Edunet> & C:/Users/hp/AppData/Local/Programs/Python/Python313/python.exe "e:/AICETE & Edunet/coffee_sales.py"
Model Evaluation:
Mean Squared Error: 0.79
R2 Score: 0.96

Model Coefficients:
```

	Feature	Coefficient
28	coffee_name_Cocoa	9.827475
27	coffee_name_Cappuccino	9.812987
32	coffee_name_Latte	9.781796
31	coffee_name_Hot Chocolate	9.753614
26	coffee_name_Americano with Milk	4.935424
25	cash_type_cash	1.323496
18	hour_16	0.570958
24	hour_22	0.455914
20	hour_18	0.347354
12	hour_10	0.346046
14	hour_12	0.290758
15	hour_13	0.288361
8	day_5	0.274694
6	day_3	0.199089
21	hour_19	0.167856
19	hour_17	0.164874
17	hour_15	0.158001
16	hour_14	0.138231
5	day_2	0.095135
7	day_4	0.089492
23	hour_21	0.074238
13	hour_11	0.062855
9	day_6	0.053696
22	hour_20	0.053190
4	day_1	0.045242
10	hour_08	-0.028731
29	coffee_name_Cortado	-0.090195
11	hour_09	-0.146980
0	month_2024-04	-0.317911
2	month_2024-06	-0.893711
1	month_2024-05	-0.924665
30	coffee_name_Espresso	-4.719472
3	month_2024-07	-5.275066

```
PS E:\AICETE & Edunet>
```

- **METRICS:**

- MEAN SQUARED ERROR (MSE): **0.79** (*LOWER IS BETTER*).
- R² SCORE: **0.96** (*HIGH ACCURACY*).

- **TOP FEATURES IMPACTING SALES:**

- COFFEE TYPES WITH THE HIGHEST DEMAND: COCOA, CAPPUCCINO, LATTE.
- PEAK HOURS INFLUENCING DEMAND.
- CASH TRANSACTIONS SLIGHTLY AFFECTING SALES.

KEY FINDINGS & BUSINESS APPLICATIONS

- OPTIMIZING STOCK FOR HIGH-DEMAND COFFEES.
- ADJUSTING RESTOCKING SCHEDULES BASED ON WEEKDAY DEMAND.
- LEVERAGING MACHINE LEARNING FOR INVENTORY PREDICTIONS.

CONCLUSION

- **PROJECT SUMMARY:**

- MACHINE LEARNING SUCCESSFULLY PREDICTS COFFEE SALES.
- HELPS OPTIMIZE VENDING MACHINE INVENTORY.

- **FUTURE SCOPE:**

- ADVANCED FORECASTING MODELS (NEURAL NETWORKS).
- CUSTOMER SEGMENTATION FOR PERSONALIZED RECOMMENDATIONS.

REFERENCES

- RESEARCH PAPERS/ARTICLES ON MACHINE LEARNING & SALES PREDICTION.
- DATA SOURCE: GITHUB

The image features a light gray background with a subtle pattern of concentric circles. In the corners, there are several realistic water droplets of various sizes, some with highlights and shadows, giving them a three-dimensional appearance. The text "Thank you" is centered in a bold, black, sans-serif font.

Thank you