VICTORIA SOLUTIONS
CONSULTING GROUP

Week 3 Project

Advanced Data Analysis Techniques and Business Insights

NO WORD COUNT

VICTORIA SOLUTIONS
CONSULTING GROUP

Objective:

To apply advanced data analysis techniques to derive actionable business insights. This includes predictive analytics, statistical modeling, and machine learning approaches for forecasting and decision-making.

VICTORIA SOLUTIONS
CONSULTING GROUP

Tools Needed

- Google Sheets or Excel (for data preprocessing and basic analysis)
- Python (Pandas, Scikit-learn, Statsmodels) for modeling
- Power BI or Tableau (optional for visualization)

The Raw Sales Data will be sent to the WhatsApp group

1. Data Preprocessing and Cleaning

Identified Issues in the Data

- Missing values in key attributes like customer demographic details.
- Outliers in the sales data affecting trend analysis.
- Inconsistent categorical variables (e.g., different labels for the same category

Steps to Follow:

- Handle missing values using appropriate imputation techniques (mean, median, mode).
- Detect and remove outliers using Z-score or IQR method.
- Standardize categorical variables for consistency.

Example Using Python:

```python
# Handling missing values
import pandas as pd

data = pd.read_csv("sales_data.csv")
data.fillna(data.mean(), inplace=True)

# Removing outliers using Z-score
from scipy.stats import zscore
data = data[(zscore(data['Sales']) < 3)]
```

2. Predictive Modeling for Sales Forecasting

Steps to Follow:

1. Apply Linear Regression to predict sales based on marketing spend and seasonality.
2. Implement Logistic Regression to classify whether a customer will churn based on historical data.
3. Use Time Series Forecasting (ARIMA/Prophet) to predict future monthly sales.

Example Using Python:

```
from sklearn.linear_model import LinearRegression
X = data[['Marketing_Spend', 'Seasonality_Index']]
y = data['Sales']
model = LinearRegression()
model.fit(X, y)
print(model.coef_, model.intercept_)
```

3. Statistical Analysis for Business Insights

Steps to Follow:

1. ANOVA: To compare sales performance across different regions.
2. Hypothesis Testing: To validate the impact of promotions on sales growth.
3. Factor Analysis: To identify key drivers influencing customer purchase decisions.

Example Using Python:

```
from scipy.stats import f_oneway
region_1 = data[data['Region'] == 'North']['Sales']
region_2 = data[data['Region'] == 'South']['Sales']
region_3 = data[data['Region'] == 'East']['Sales']

anova_result = f_oneway(region_1, region_2, region_3)
print("ANOVA P-value:", anova_result.pvalue)
```

4. Machine Learning for Customer Segmentation

Steps to Follow:

- Use Decision Trees to segment customers based on purchasing behavior.
- Implement K-Means Clustering to group customers into different spending categories.
- Apply Ensemble Learning (Random Forest, XGBoost) for enhanced prediction accuracy.

Example Using Python:

```
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=3)
data['Customer_Segment'] = kmeans.fit_predict(data[['Total_Spend', 'Purchase_Frequency']])
```

5. Business Insights & Recommendations

Key Findings:

- High-Value Customers: Identified through clustering; targeted offers should be provided.
- Sales Forecasting: Predictive models indicate seasonal spikes, allowing inventory optimization.
- Churn Prevention: Logistic regression helps in identifying at-risk customers early.

Recommended Business Actions:

- Personalize marketing strategies based on customer segmentation results.
- Adjust stock levels based on time series forecasting to avoid overstocking or shortages.
- Implement customer retention programs for segments with high churn probability.

VICTORIA SOLUTIONS
CONSULTING GROUP

Final Deliverables

Submit your work in the project coversheet containing:

1. Cleaned dataset (Google Sheets or Excel)
2. Python scripts for predictive modeling and statistical analysis
3. Three visualizations (e.g., time series forecast, customer segmentation, ANOVA results)
4. Summary report detailing key findings and business recommendations

VICTORIA SOLUTIONS
CONSULTING GROUP

Please write the answers in the “Project Coversheet” which is shared in the WhatsApp group and for completing the tasks refer to the Data Set provided in the group