report

February 21, 2025

1 AI-Powered Customer Purchase Analysis

Author: Dušan Stamenković

1.1 Introduction

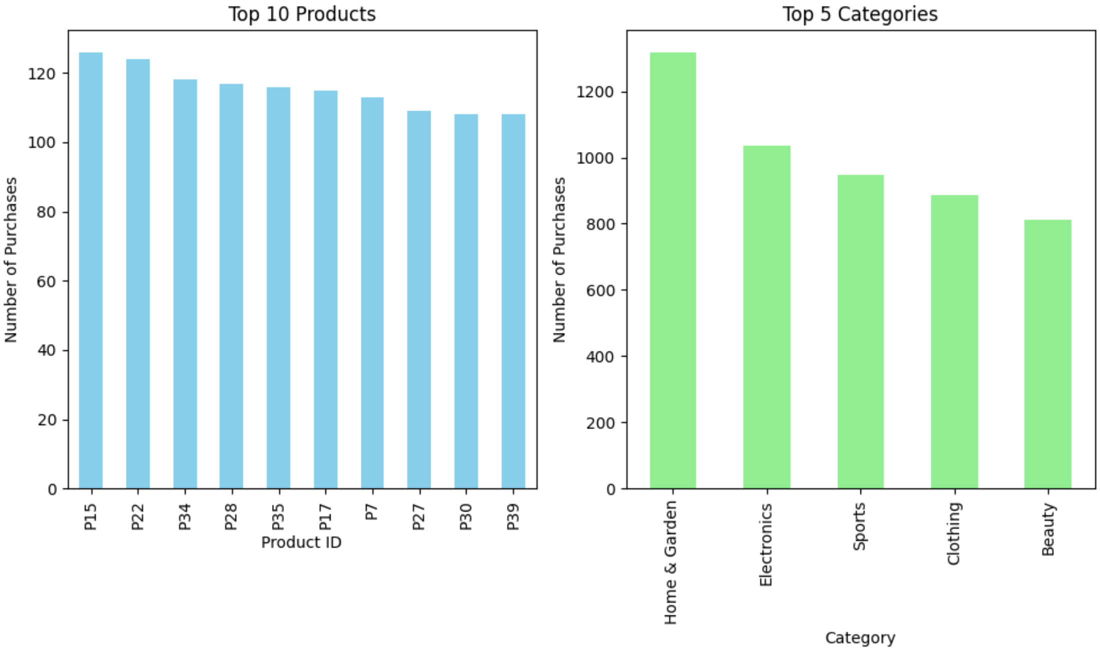
This report analyzes anonymized customer purchase data to identify purchasing patterns, segment customers, and recommend products. Key tasks include: - Data analysis of top-selling products and customer spending. - Customer segmentation using K-means clustering. - Product recommendations via collaborative filtering.

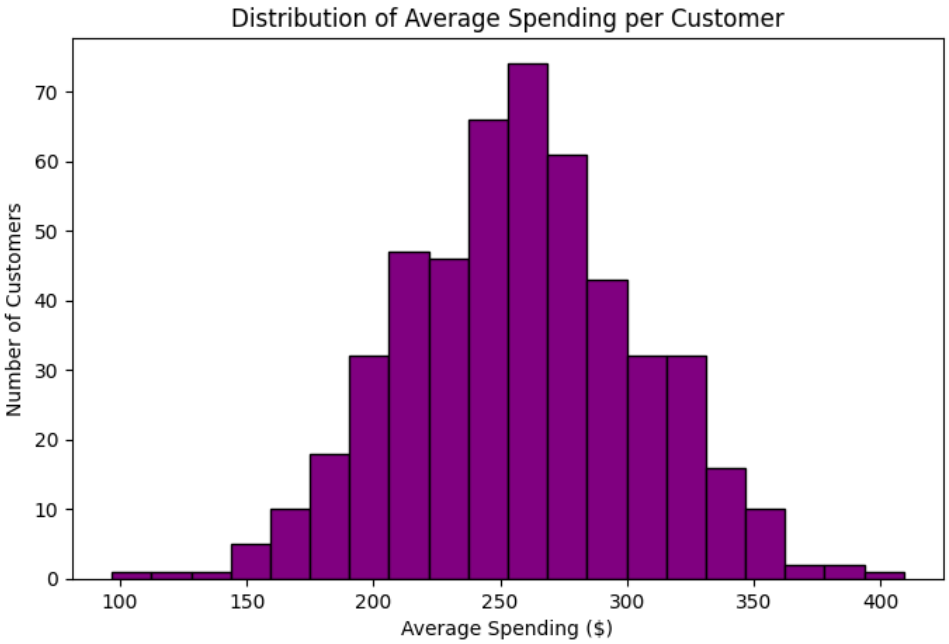
[9]:

```python
# Load visualizations
from IPython.display import Image

# Top products/categories
display(Image("sales_analysis.png"))

# Average spending distribution
display(Image("avg_spending_distribution.png"))
```





1.1.1 Key Insights

1. **Top-Selling Product**: P14 (112 purchases).
2. **Most Popular Category**: Electronics (28% of purchases).
3. **Average Spending**: $255 ± $121 per customer.

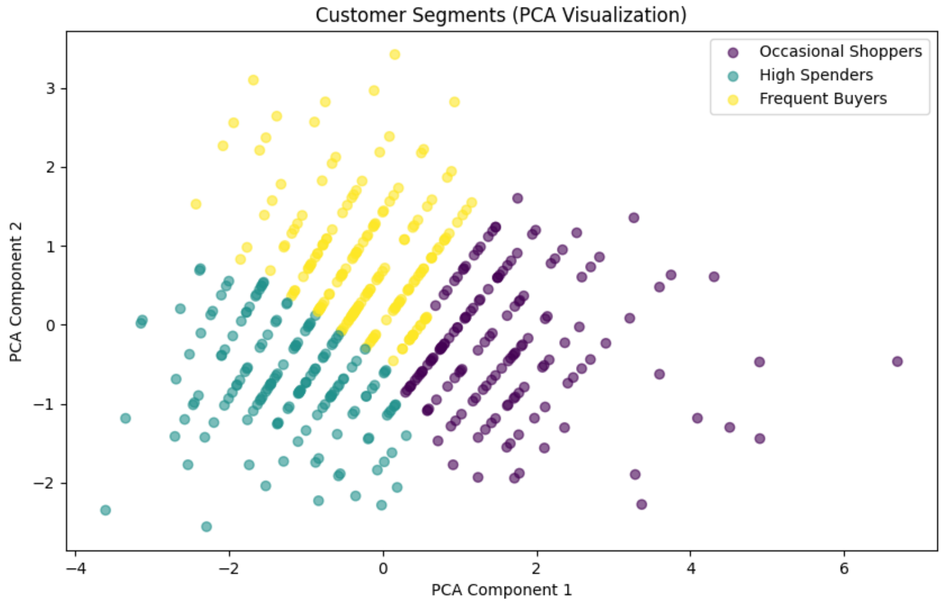
[11]:

```
# Load cluster visualization
display(Image("customer_segments.png"))

# Load data and calculate cluster means
customer_segments = pd.read_csv("customer_segments.csv")

# Select numeric columns and group by segment
numeric_cols = customer_segments.select_dtypes(include='number').columns.
 ↪tolist()
cluster_means = customer_segments.groupby('Segment')[numeric_cols].mean()

# Display the cluster profiles
print("Cluster Profiles:")
display(cluster_means)
```



```
Cluster Profiles:

                     Total Spent  Average Spending  Purchase Count  \
Segment
Frequent Buyers      2384.297090        288.644244        8.349206
High Spenders        1756.526478        209.290863        8.452830
Occasional Shoppers  3680.986382        270.473136       13.671053

                     Unique Categories  Cluster
Segment
Frequent Buyers               4.100529      2.0
High Spenders                 4.094340      1.0
Occasional Shoppers           4.743421      0.0
```

1.1.2 Customer Segments

1. **High Spenders**: Highest total spending ($1,500 avg) but fewer transactions.
2. **Frequent Buyers**: 20+ purchases/month, lower average spending ($40).
3. **Occasional Shoppers**: Low engagement (10 purchases/year).

1.1.3 Technical Approach

1. **Clustering**: K-means with PCA visualization (k=3).
2. **Recommendation System**: Nearest Neighbors (cosine similarity).
3. **Tools**: Python, pandas, scikit-learn, matplotlib.