

Exploratory Data Analysis on Customer Behavior

Internship Project – Internet Brands, September 2023

Project Overview

As part of my internship, I conducted exploratory data analysis (EDA) to uncover trends in customer behavior across 3,500 markets. The goal was to guide pricing strategy updates, identify underperforming markets, and optimize resource allocation.

The analysis revealed significant patterns, leading to the removal of 33% of underperforming markets and the development of a streamlined pricing model. These changes improved profitability and operational efficiency, demonstrating the impact of data-driven decision-making.

Data Collection and Preparation

1. Data Sources:

- Customer interaction data from platforms like Avvo and Martindale-Hubbell, including:
 - Monthly ad spend.
 - Call volumes and conversion rates.
 - Retention and churn rates.
- External market data such as population density, competition levels, and regional economic indicators.

2. Data Cleaning and Wrangling (Python – Pandas):

- Handled missing values using imputation techniques based on market averages.
- Consolidated data from multiple sources into a unified DataFrame.

Standardized numerical fields for accurate trend comparison, e.g.:

python

```
# Normalizing ad spend by market size  
df['normalized_spend'] = df['ad_spend'] / df['market_population']
```

Exploratory Data Analysis Process

1. Univariate and Bivariate Analysis (Pandas, Matplotlib):

Customer Spend Distribution:

Analyzed the distribution of ad spend across markets to identify clusters of high and low spenders.

python

Copy code

```
# Visualizing ad spend distribution
df['ad_spend'].plot(kind='hist', bins=50, title='Ad Spend
Distribution')
plt.xlabel('Ad Spend ($)')
plt.ylabel('Frequency')
plt.show()
```

- **Calls-to-Spend Ratio (CPD):**
Investigated the relationship between calls and spend to flag underperforming markets with high spend but low call volume.
 - 2. **Segmentation Analysis:**
 - Categorized markets into performance tiers (e.g., High, Moderate, Low) based on conversion rates and CPD metrics.
 - Identified regional trends in performance, such as urban markets outperforming rural ones.
 - 3. **Trend Identification:**
 - Detected seasonal variations in customer behavior, which were later integrated into pricing recommendations.
 - Visualized churn rates over time to identify patterns linked to pricing changes or external market factors.
-

Insights and Findings

1. **Underperforming Markets:**
 - 33% of markets exhibited low conversion rates and high churn, making them candidates for resource reallocation or exit.
 2. **Optimal Pricing Strategies:**
 - Developed a streamlined pricing model by analyzing CPD trends and competitor benchmarks, ensuring alignment with market demand.
 3. **Customer Retention Drivers:**
 - High customer engagement correlated with markets offering personalized ad strategies, suggesting opportunities for targeted campaigns.
-

Visualization and Reporting

1. **Summary Dashboard (Matplotlib/Seaborn):**
 - Created visual summaries, including:
 - Heatmaps of CPD performance across regions.

- Line plots for churn and conversion trends.

2. Stakeholder Presentation:

- Delivered actionable insights with clear visualizations, highlighting the business impact of market exits and pricing updates.
 - Provided a detailed report summarizing key trends and recommendations.
-

Key Outcomes

- Identified and removed 33% of underperforming markets, freeing resources for higher-growth areas.
- Informed the creation of a pricing model that improved revenue predictability and operational efficiency.
- Empowered stakeholders to make data-driven decisions by presenting clear, impactful findings.

This project highlights my expertise in Python-based EDA, ability to derive actionable insights, and effectiveness in communicating complex data to drive business outcomes.