

ANALYTICS FOR A NEW COFFEE BRAND LOOKING TO ENTER THE USA MARKET

Sena Dilan Çakır - 21050911027 Musa Yüksel - 21050911018

ABSTRACT

This study aims to identify the most suitable market entry cities for a new premium coffee brand in the United States. The study has developed a data science-based quantitative analysis framework. Demographic indicators, competition intensity, and consumer engagement were combined to create an Opportunity Index at the city level. The analysis shows that market opportunities are not homogeneous across the United States and that only a limited number of cities offer high potential.

RESEARCH QUESTIONS

- Which cities in the US are most suitable for a new premium coffee brand?
- How do demographics, competition, and consumer behavior affect market opportunities?
- Is competition or demand more decisive in market entry?

DATASET DESCRIPTION

- U.S. Census Bureau (ACS 2022)
 - Population
 - Median household income
 - Median age
Census County Business Patterns (CBP)
 - NAICS 72251
 - Number of businesses by county
Yelp Fusion API
 - Average rating
 - Number of reviews
 - Percentage of premium-priced cafes
Department of Transportation (DOT) Crosswalk
 - City-county FIPS mapping

DATASET PREPROCESSING

- Most plausible cities for a cafe has been filtered from Census ACS using filters
 - Population: > 150,000
 - Median Household Income: > \$55,000
 - Median Age: 22-45 years old
 - City names have been standardized
 - FIPS codes have been cleaned and normalized
 - City-based demographic data has been matched with county-based competitive data
 - Incomplete and erroneous records have been excluded from the analysis
- This process has resulted in a unified analysis dataset with high accuracy.

KEY FINDINGS

Market opportunity is determined by the balance between demand and competition rather than by population or income alone. Cities with strong demand and manageable competition offer the most attractive entry points, while highly competitive metropolitan markets require greater differentiation and investment. NAICS-based cafe density provides a more reliable view of competition than online listings, and clustering results show that cities with similar opportunity scores can represent different market types, highlighting the need for tailored entry strategies.

INDEX CALCULATIONS

Indices calculates with feature effectiveness weights Demand Index:

- Median income (%40)
- Population (%30)
- Yelp engagement level (%30)

Competition Index:

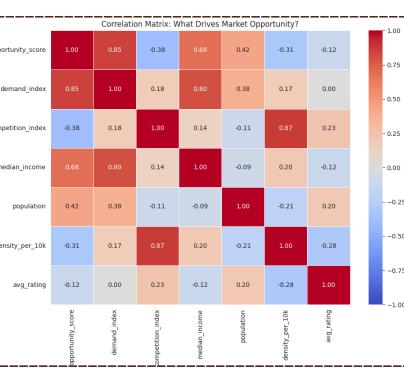
- Cafe density (70%)
- Average rating (30%)

Opportunity Index:

$$\text{Opportunity} = \text{Demand} - (0.6 \times \text{Competition})$$

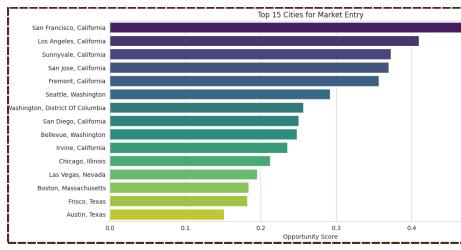
FACTORS MOST STRONGLY INFLUENCE OPPURTUNITY SCORE

We observe a strong positive correlation between median income and opportunity score. There is a negative correlation between cafe density and opportunity score. Engagement often correlates positively with opportunity, while avg rating has a weaker impact.



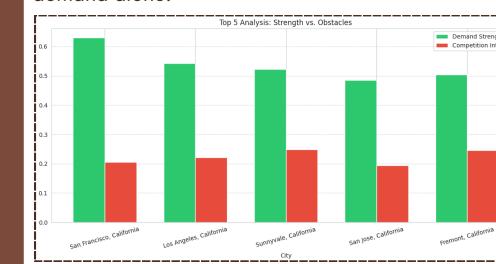
TOP 15 CITIES

Based on the **Opportunity Index**, we have identified the top 15 cities for expansion. These markets exhibit a unique combination of high disposable income and an active coffee culture, yet lack the oversaturation. Based on our calculations the most promising plausible place for a market entry is **San Francisco, California**



DEMAND VS. COMPETITION

The analysis highlights that the most attractive entry opportunities arise where strong consumer demand is paired with manageable competition rather than demand alone.

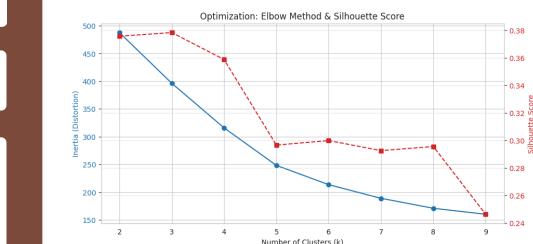


UNSUPERVISED MARKET SEGMENTATION WITH MACHINE LEARNING

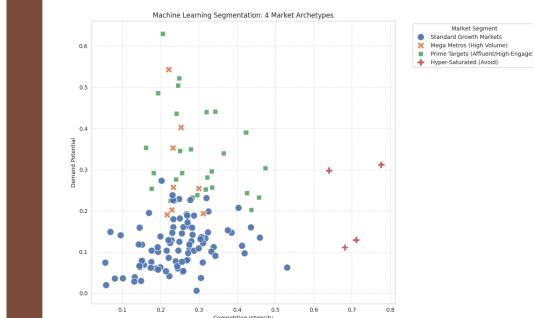
We used **K-Means Clustering** to discover latent market archetypes.

Methodology:

- We select normalized indicators of Demand, Competition, and Affordability. We exclude the 'Opportunity Score' itself to avoid circular logic.
- Data is standardized so that variables with large ranges don't dominate variables with small ranges.
- We use the "Elbow Method" and "Silhouette Score" to determine the mathematically optimal number of clusters.



- We assign human-readable names to each cluster based on their centroids.



- Green Points (Prime Targets): Cities with high demand, low competition.
- Red Points (Hyper-Saturated): These outliers are pushed far to the right, visually confirming their extreme density.
- Orange/Blue Points: The distinct separation between "Standard" (Blue) and "Mega Metros" (Orange) highlights that while both have moderate opportunity scores, they require fundamentally different operational scales.

DISTRIBUTION OF OPPURTUNITY

Most cities cluster around slightly negative to modestly positive scores, indicating that average market conditions are competitive and do not inherently favor new entry.



CONCLUSION

This study reveals that opportunities are not evenly distributed in the US coffee market. The developed Opportunity Index provides a powerful tool for data-driven market entry decisions. The method can also be easily adapted to different retail sectors.

REFERENCES

- U.S. Census Bureau. American Community Survey (ACS) 5-Year Estimates. <https://www.census.gov/programs-surveys/acs>
- U.S. Census Bureau. County Business Patterns (CBP)(NAICS category 72251 – Restaurants and Similar Eating Places). <https://www.census.gov/programs-surveys/cbp>
- Yelp Inc. Yelp Fusion API. <https://docs.developer.yelp.com/docs/fusion-intro>
- North American Industry Classification System (NAICS). <https://www.census.gov/naics>