IronHack Data Analytics

WEEK 8+9 | FOOD DELIVERY INSIGHTS

Agenda

- Datasource & Availability
- Business Case
- Hypothesis
- Database Design
- Data Insights
- Conclusions & Business Implications

Data Source & Availability

- Data sourced via Kaggle
  - Dataset contains **lists of restaurants** and **their menus** in the **US** that are partnered with Uber Eats
  - Data **collected via web scraping** using Python libraries

restaurants   restaurant-menus

Business Case

Help restaurant owners, investors, and platform operators optimize operations and improve menu offerings by providing insights into restaurant performance, menu pricing, and customer preferences

Key Business Questions

1. **Restaurant Performance & Ranking:**

- What factors contribute to higher restaurant rankings?
- How do ratings correlate with menu prices and restaurant categories?
- Are there geographic trends in restaurant ratings and rankings?

2. **Menu Insights & Pricing Strategy:**

- What are the average price ranges for different cuisines?
- Which menu categories contribute most to high-rated restaurants?
- Are there pricing trends across locations or restaurant types?

3. **Market Trends & Consumer Preferences:**

- What are the most common cuisines and menu categories?
- Which restaurant categories tend to have the best ratings?
- Are certain cuisines more popular in specific locations?

4. **Geospatial Analysis:**

- Which areas have the highest-rated restaurants?
- How does restaurant density vary by location?
- Are higher-priced restaurants concentrated in specific neighborhoods?

Hypothesis

Hypothesis 1: **The most common cuisines tend also to be the most affordable.**

- Common cuisines indicate stronger customer preferences and affordability plays a critical part in the selection process

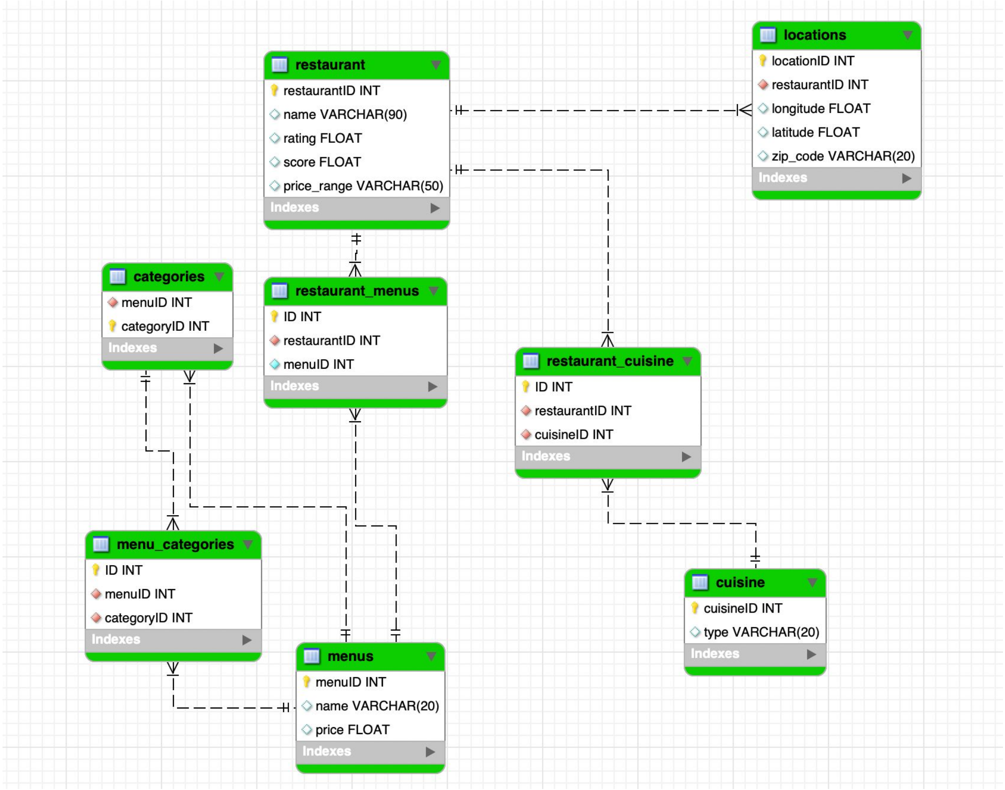
Hypothesis 2: **Most restaurants either have extremely low or extremely high ratings, with few in the middle.**

- It's either restaurants which just started out recently or those that were able to grow their business on Uber Eats long term – mediocre restaurants won't be able to survive

Data Cleaning

restaurants	restaurant-menus
Fixed Price Range Column: ● Transformed price range symbols: ○ $ → "Cheap" ○ $$ → "Moderate" ○ $$$ → "Expensive" ○ $$$$ → "Very Expensive" **Standardized Categories:** ● Identified top categories and cleaned category names to reduce redundancy. ● Mapped multiple-word categories to broader terms (e.g., "Burgers, American, Sandwiches" → "Burgers"). **Verified Data Types:** ● Ensured id remained an integer, while categorical values were kept as strings.	**Converted Price Column to Float:** ● Ensured all values in price were numeric for analysis. **Trimmed and Cleaned Text Columns:** ● Standardized casing and removed unnecessary whitespace. **Linked with Restaurants Dataset:** ● Ensured restaurant_id in restaurant-menus matched valid id values in restaurants.

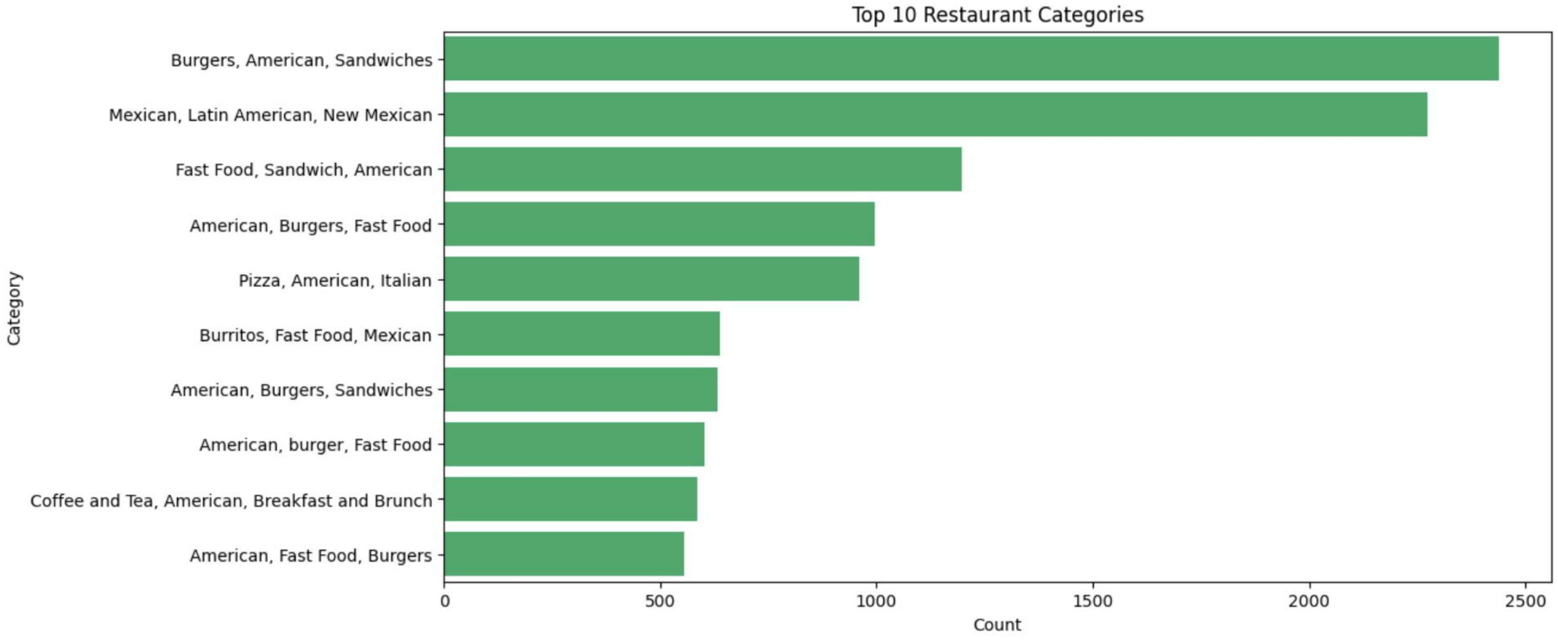
Database Design



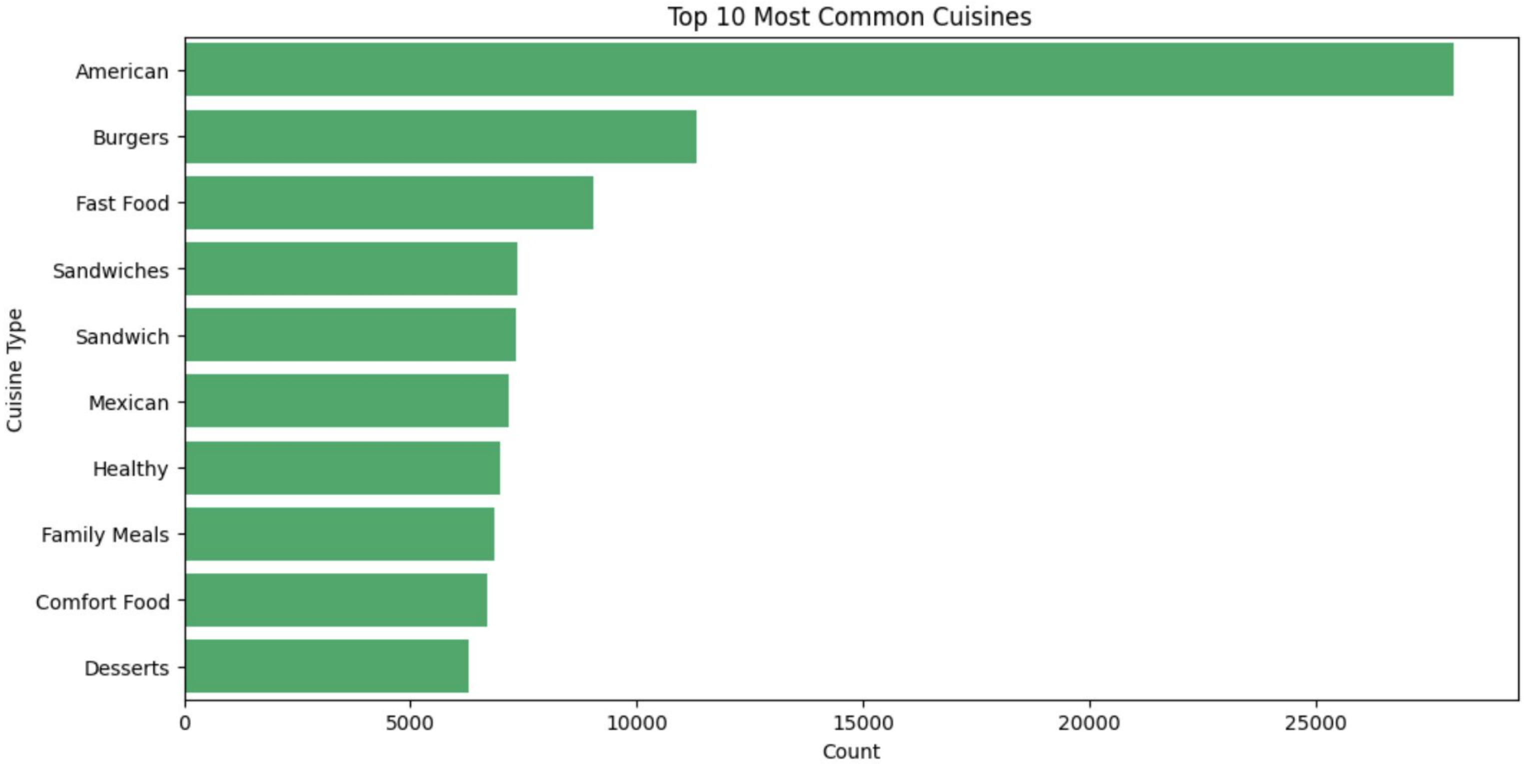
Key Challenges:

- Handling many-to-many relationships
- Cleaning & normalizing data
- Optimizing query performance

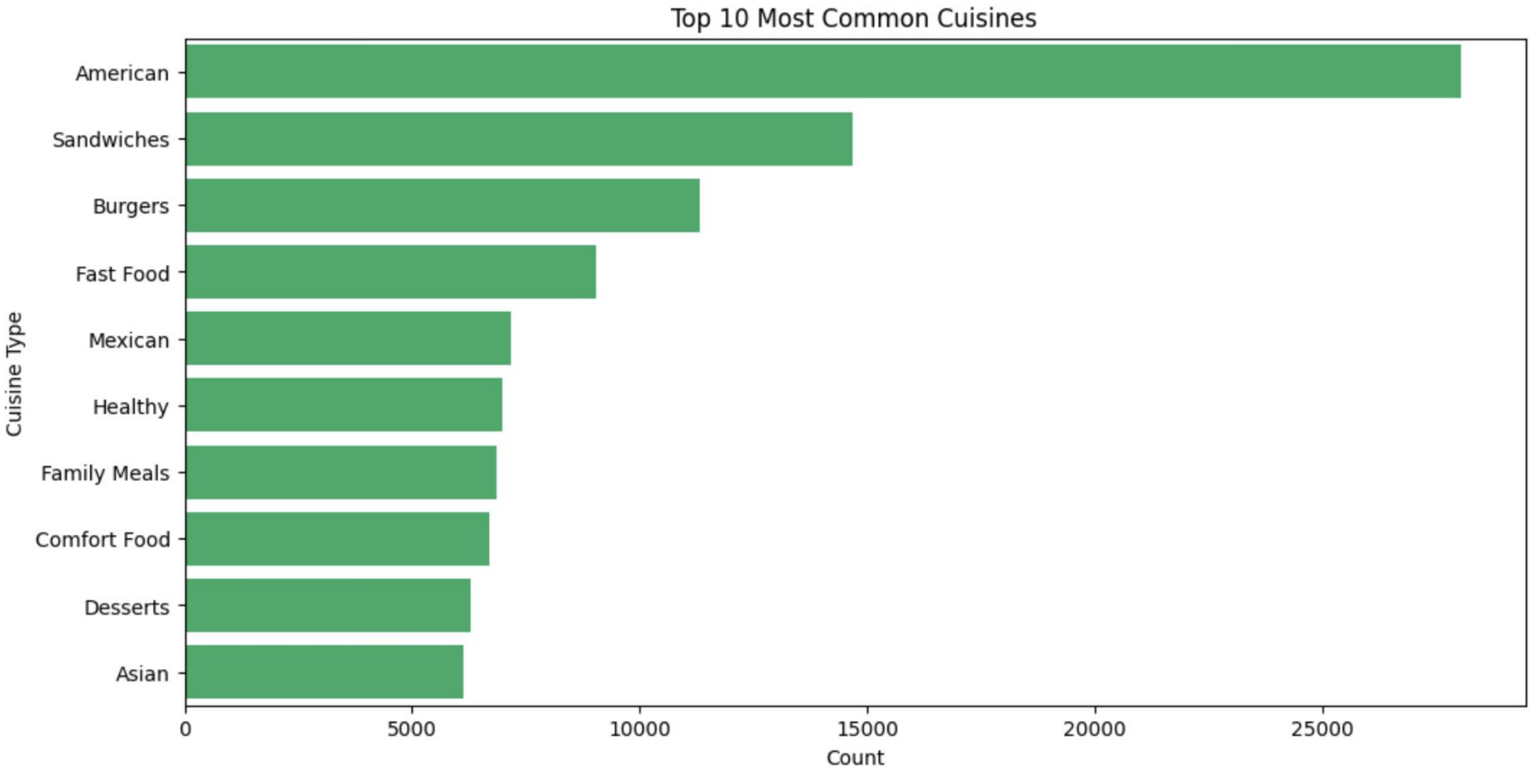
Top 10 Restaurant Categories



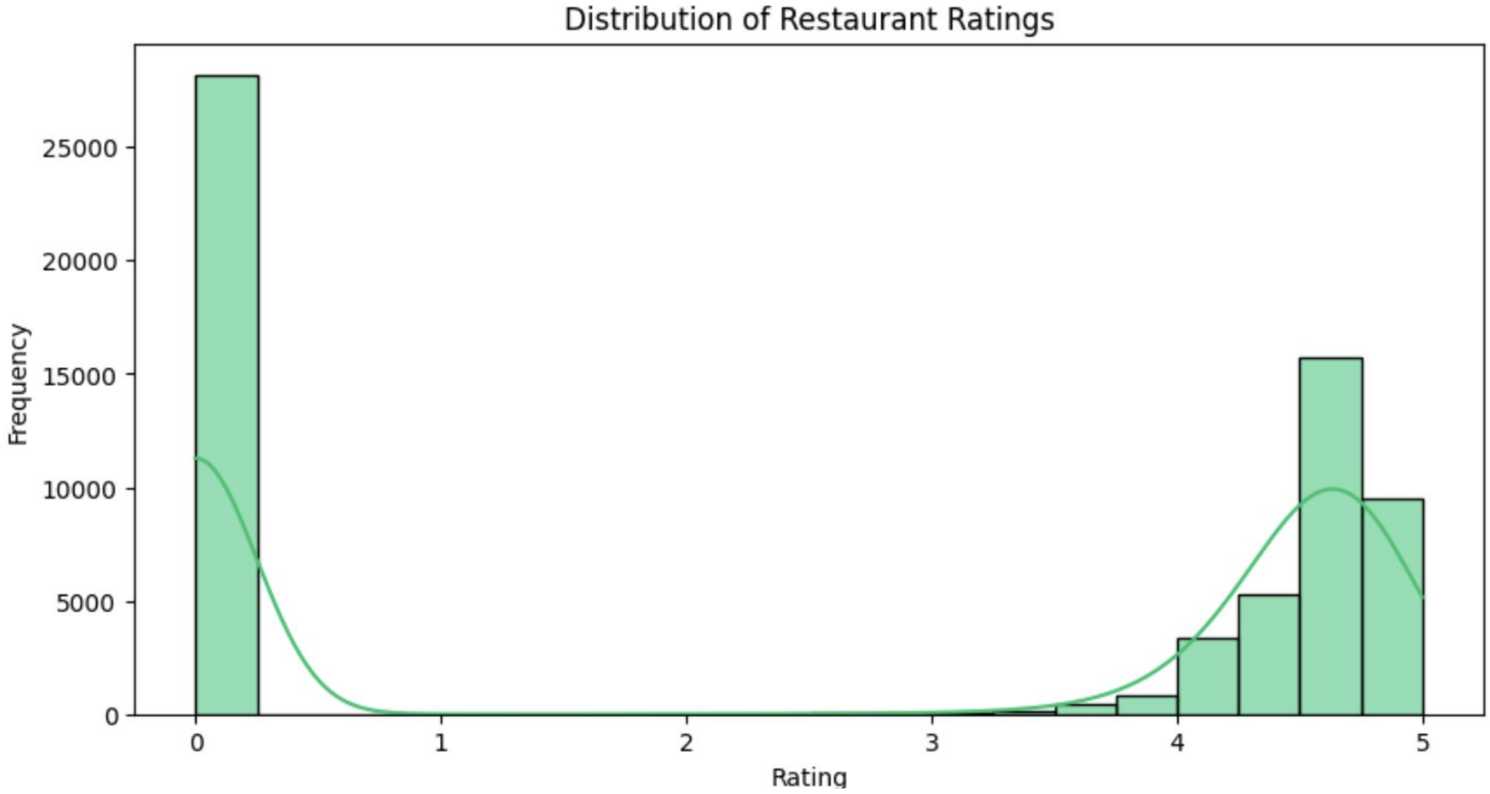
Restaurant Performance & Ranking



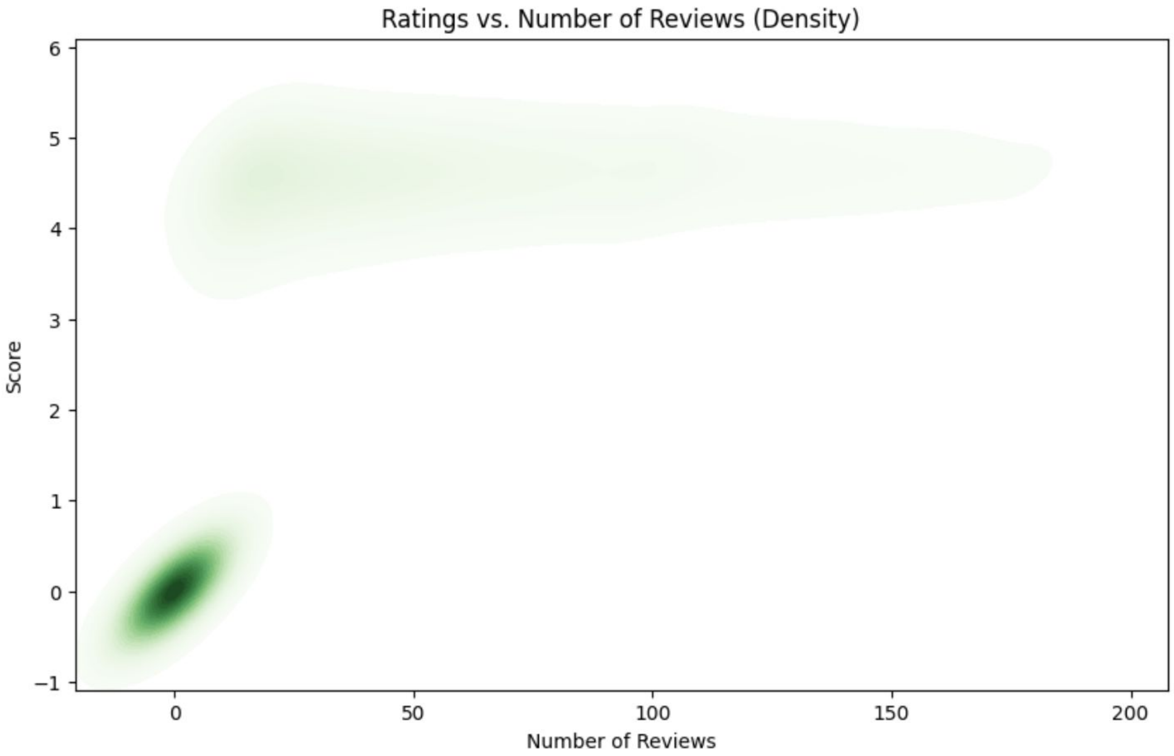
Top 10 Most Common Cuisines



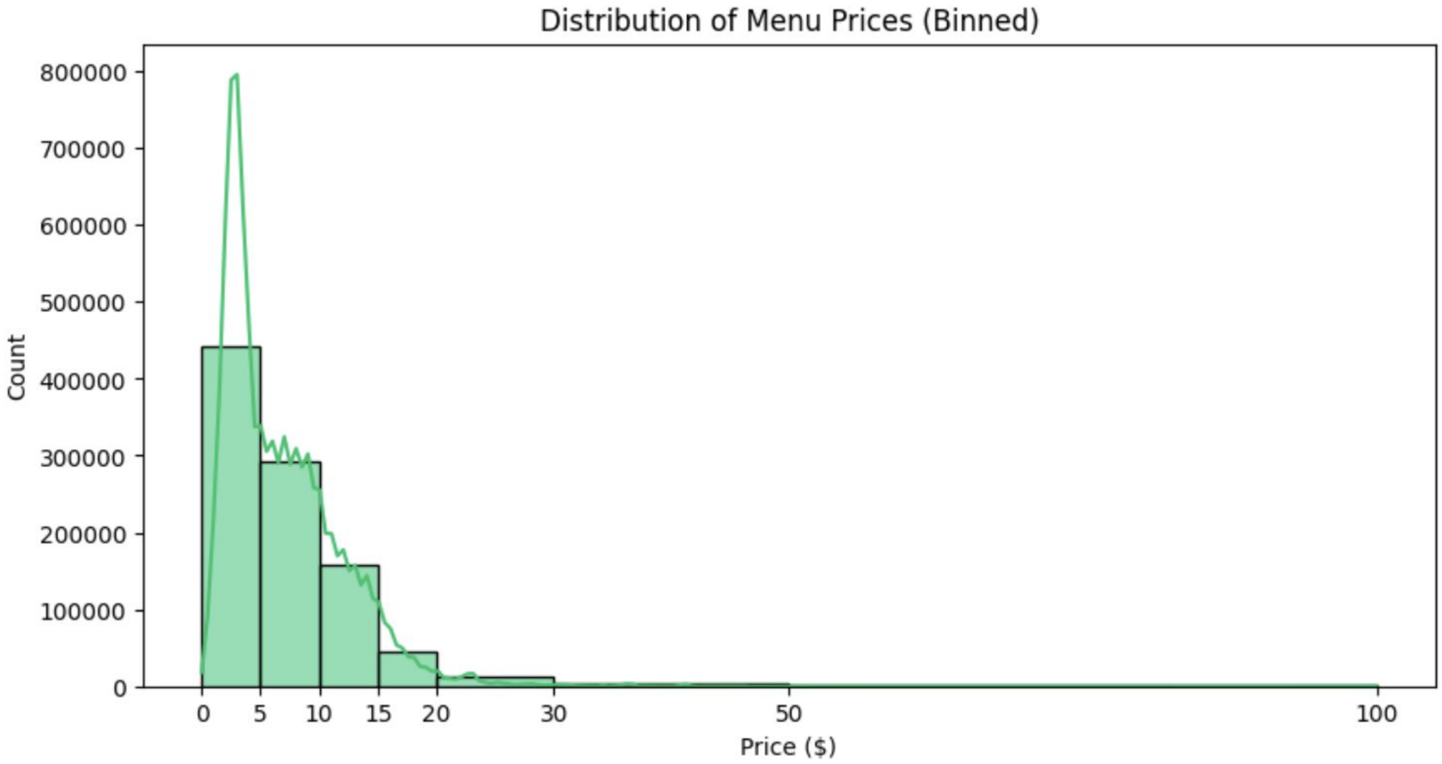
Restaurant Ratings by Frequency



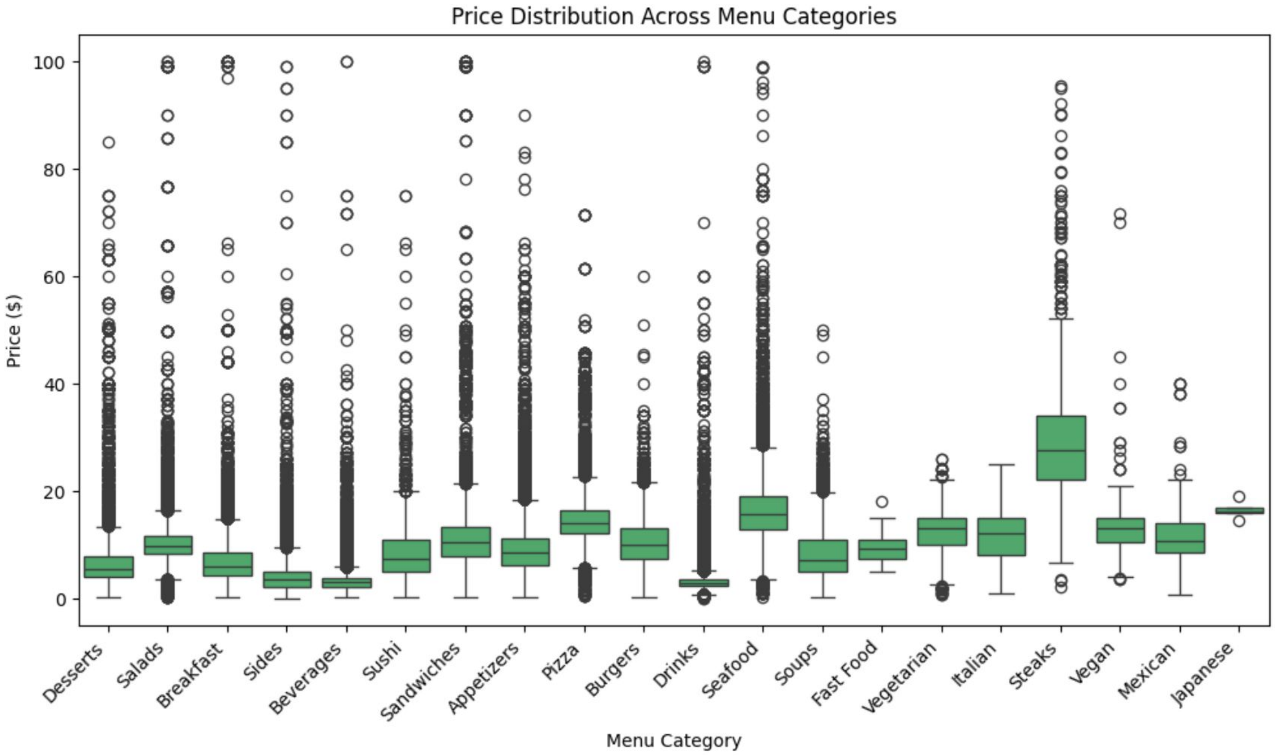
Restaurant Ratings vs. Number of Reviews



Price Distribution of Menu Items



Price Distribution across Menu Categories

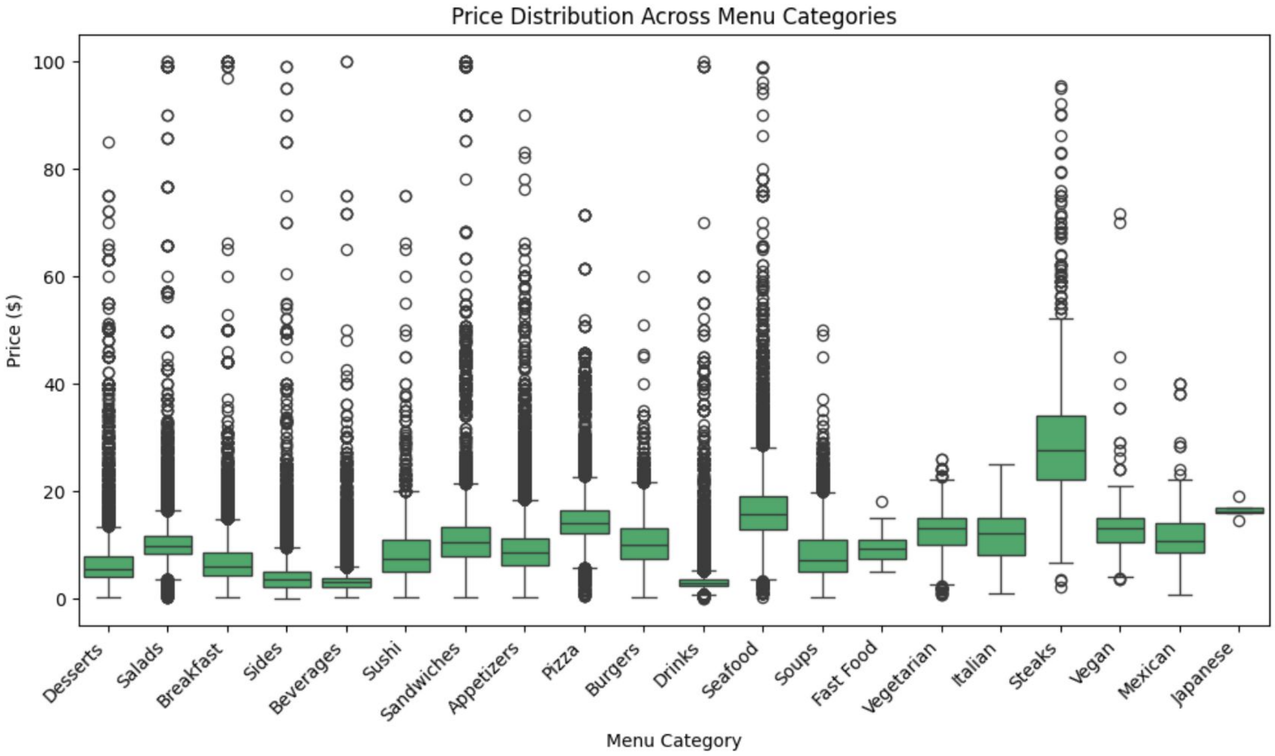


Conclusions & Business Implications

Hypothesis 1: **The most common cuisines also tend to be the most affordable.**

- If affordability drove popularity, we would expect common cuisines to be cheaper
- **Data shows:** Some popular cuisines have a **wide price range**
- **Takeaway:** Factors like **brand recognition & convenience** may drive demand more than price

Price Distribution across Menu Categories



Conclusions & Business Implications

Hypothesis 2: **Most restaurants either have extremely low or extremely high ratings, with few in the middle.**

- If quality was evenly spread, ratings should cluster around **3 stars**
- **Data shows:** Strong **bimodal distribution** – either **0 stars or 4.5-5 stars**
- **Takeaway:** Restaurants tend to **either thrive or struggle**, with few truly "average" places

Major Obstacles

- Struggled with **outliers** & **data distribution**: Menu prices had extreme outliers (e.g., some dishes were listed at $1000+)

- **Category** column in menus was a **nightmare:** Categories were **highly inconsistent** and bunched up (e.g., 'Burgers, American, Sandwiches' vs. 'American, Burgers, Fast Food')

- **SQL** database **setup** was a **pain**: Creating a SQL database for this dataset was more time-consuming and complex than expected

Learnings

- **Deep dive** on **data cleaning**: Handling duplicate and inconsistent categories (e.g., ‘Sandwich’ vs. ‘Sandwiches’) of a seemingly “clean” data set is critical to extracting insightful takeaways

- **Focus** on the **relevant data**: Keep analysis centered around hypothesis

- The **right visualization** makes a **huge difference**: Some scatter plots were unreadable at first, but switching to KDE plots made the insights clearer

FOOD DELIVERY INSIGHTS

IRON HACK

THANKS !

Lukas Günther