Restaurant Analysis Project Plan

1. Project Overview

This project explores the key factors that influence restaurant success, focusing on popularity, customer ratings, order volume, and customer demographics. By combining insights from three datasets—`restaurant.csv`, `orders.csv`, and `users.csv`—we aim to identify what drives customers to highly rate certain restaurants, whether these ratings translate to higher order volumes, and how user demographics play a role in shaping perceptions and purchasing behavior.

The project leverages data transformation, relational database queries, and data visualization tools to extract actionable insights that can guide restaurant strategy and marketing decisions.

---

2. Research Questions

Primary Research Question

- **Which factors most influence customer ratings and restaurant popularity, and do these same factors drive higher order volume and revenue?**

Supporting Questions

- Do cuisine type, cost level, and geographic location correlate with higher ratings or more frequent customer orders?

- Are the most highly rated restaurants also the most financially successful?

- How do customer demographics (age, income, family size, etc.) influence their ratings, preferences, and spending behavior?

---

3. Hypothesis

We hypothesize that:

- **Cuisine type** and **geographic location** are the most significant factors in determining a restaurant's popularity and customer rating.

- **Cost level** has a minimal influence on customer satisfaction compared to the type of cuisine and convenience of location.

- **Customer demographics**, including income and age, influence both restaurant preference and rating behavior, with variation across cuisine and pricing categories.

---

4. Datasets

A. `restaurant.csv`

Purpose: Core restaurant data
 Key Fields:

- `id`: Unique restaurant identifier

- `name`: Restaurant name

- `city`, `address`: Location

- `rating`: Average customer rating

- `rating_count`: Number of reviews

- `cost`: Price tier (e.g., $, $$, $$$)

- `cuisine`: Type of cuisine

B. `orders.csv`

Purpose: Order-level transaction data
 Key Fields:

- `r_id` (to be renamed to `restaurant_id`): Foreign key to `restaurant.csv`

- `order_date`: Date of order

- `user_id`: Foreign key to `users.csv`

- `sales_qt`: Quantity of items sold

- `sales_amount`: Total value of the order

C. `users.csv`

Purpose: Customer demographic data
Key Fields:

- `user_id`: Unique user identifier

- `name`: Customer name

- `age`: Age

- `gender`: Gender identity

- `marital status`: Marital status

- `family size`: Number of household members

- `monthly income`: User income

---

5. Tools and Technologies

Tool	Purpose
Google Sheets	Data cleaning, normalization, and light aggregation
SQL (PostgreSQL / MySQL / SQLite)	Joins, aggregation, filtering, and advanced querying
Tableau	Interactive dashboards and visual storytelling

6. Data Integration Strategy

A. Key Relationships

To enable holistic analysis across restaurants, transactions, and customer profiles, we will create the following joins:

1. `restaurant.csv` → `orders.csv`
   Join condition: `restaurant.id = orders.restaurant_id`
   (Note: rename `r_id` to `restaurant_id` for consistency)

2. `orders.csv` → `users.csv`
   Join condition: `orders.user_id = users.user_id`

These joins will be performed both:

- In SQL for analysis and reporting

- In Tableau for dashboard development and visual exploration

---

7. Data Cleaning and Preparation

Tool: Google Sheets
Purpose: Ensure all datasets are correctly formatted, standardized, and ready for analysis.

Key Cleaning Tasks:

For `restaurant.csv`:

- Standardize `cost` to numeric tiers

- Clean up the `cuisine`, `city`, and `address` text fields

- Check for missing values and duplicates

- Create a derived field for popularity:
  `Popularity Score = rating * LOG(rating_count)`

For `orders.csv`:

- Rename `r_id → restaurant_id` for consistency with `restaurant.csv`

- Format `order_date` into a date-friendly format if needed

- Ensure `sales_qt` and `sales_amount` are numeric

For `users.csv`:

- **Rename columns for clarity and consistency:**

Original Column	New Column Name
marital status	marital_status
family size	family_size
monthly income	monthly_income

- Ensure all text values are standardized (e.g., lowercase for gender)

- Validate that `user_id` is unique and matches with records in `orders.csv`

- Ensure `age`, `family_size`, and `monthly_income` are numeric

- Remove or flag invalid entries (e.g., negative income)

---

8. Analysis Plan

A. Restaurant-Level Analysis

- Which restaurants have the highest ratings and review counts?

- Which restaurants generate the highest order volume and revenue?

- How do these metrics vary by:

    - Cuisine

    - City

    - Cost level

B. User Behavior Analysis

- What are the typical spending patterns by age, income, and family size?

- Do specific demographics consistently give higher or lower ratings?

- Which cuisines are favored by different age or income groups?

C. Factor Impact Assessment

- How strongly do cost, location, and cuisine predict customer rating?

- Is there a correlation between rating and sales volume?

- Are expensive restaurants rated higher, or are budget-friendly spots more popular?

---

9. Tableau Dashboards

Dashboard Components:

1. **Restaurant Performance Overview**

    - Ratings, reviews, orders, and sales by restaurant

2. **Cuisine & City Insights**

- Cuisine types and cities with the highest ratings and sales figures

3. **Cost vs. Popularity**

    ○ Cost tiers and their relationship to ratings, review volume, and order activity

4. **User Demographics**

    ○ Preferences and spending habits segmented by age, income, gender, and family size

Each dashboard will include filters and interactivity to explore results by cuisine, cost, location, and demographic groups.

---

10. Deliverables

- Cleaned and standardized datasets (`restaurant.csv`, `orders.csv`, `users.csv`)

- SQL queries for core analysis, including joins and aggregations

- Tableau dashboards providing visual summaries of key findings

- Final written report or presentation summarizing:

    ○ Key insights from the analysis

    ○ Recommendations for improving restaurant performance

    ○ Observed trends in customer behavior and engagement

---

11. Next Steps

1. **Data Cleaning** – Complete cleanup of all three datasets in Google Sheets

2. **Database Setup** – Load data into a relational SQL environment

3. **Join and Query** – Perform key joins and analysis using SQL

4. **Data Modeling in Tableau** – Replicate joins and create visual dashboards

5. **Insight Compilation** – Write report and/or prepare slides for presentation