**Group Members**:

Denise McKinley

Tessy Azogu

Sunil Joshi

Ninglian (Melanie) Wang

Martin

**Scenario**:

Entity X (Mexican, Tex Mex, Bar) is looking to open 10 new locations in the United States. They would like to start within 1 state, but need to know which state based on consumer spending patterns and market saturation. What are the top 5 state locations based on low market saturation, high ratings and high consumer spending (within the Mexican, Tex Mex, Bar categories)?

**Null Hypothesis**:

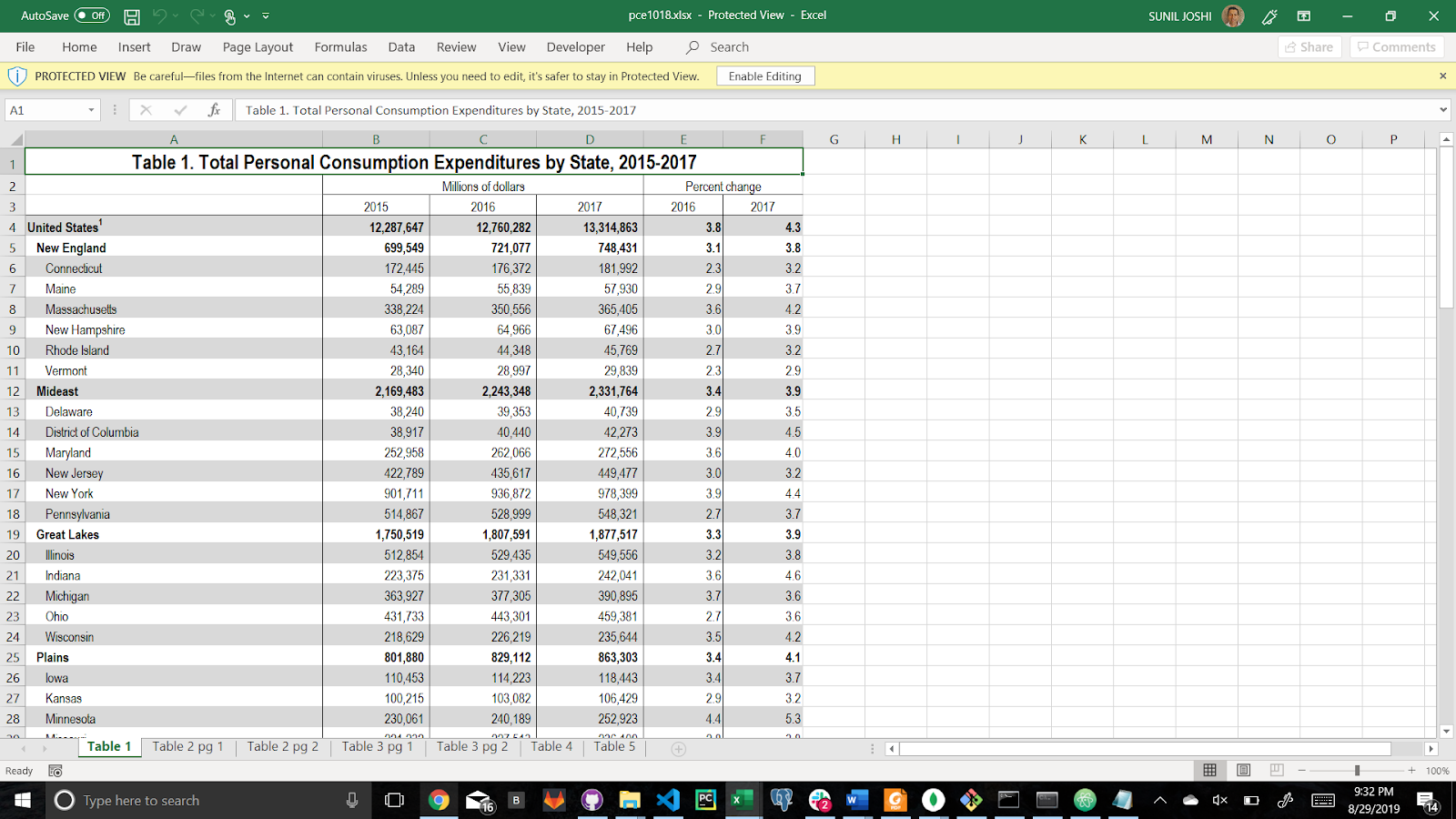
There is no difference between the 50 states in market saturation, ratings (average and spread) and consumer spending.

**Data Sources**:

BEA Personal Consumer Spending and Income by State(2019)

50 rows x

<https://www.bea.gov/data/income-saving/personal-income-by-state>

<https://www.bea.gov/data/consumer-spending/state>

Yelp Business Dataset ()

<https://www.kaggle.com/yelp-dataset/yelp-dataset/version/6#yelp_business.csv>

192,000 Rows x 13 Columns

**Things to Consider**:

* What kind of business is our client (Mexican, Tex Mex, Bar)
* Saturation these business types
* Based on rating of businesses (3-5)
* Is there a correlation between consumer spending and ratings (or any of the variables we are using to compare the states)?

**Process/Dependencies**:

### SWAT, PEST analysis

<https://creately.com/blog/diagrams/swot-analysis-vs-pest-analysis/>

<https://www.ibisworld.com/industry-trends/market-research-reports/accommodation-food-services/chain-restaurants.html>

<https://www.linkedin.com/pulse/difference-between-swot-pest-steep-steeple-analysis-mohammadpourfard/>

<https://restaurant.eatapp.co/blog/restaurant-swot-analysis>

# Narrative

As part of SWOT, Opportunity analysis customer requests solution which will provide relevant dataset to carry out the analysis.

Provided ETL must support multiyear usage, previously loaded data must be preserved.

## Demand determinants

Demand for full-service restaurants is driven by several factors, including household income, consumer confidence, attitudes to health and propensity to e…

**Customer segments**

* Households earning $100,000 to $199,999
* Households earning over $200,000
* Households earning less than $30,000
* Households earning $30,000 to $49,999
* Businesses
* Government

**Key External Drivers**

* Consumer spending
* Consumer Confidence Index
* Healthy eating index
* Households earning more than $100,000

**Business trend with impact on restaurant business**

* Health
* Supply
* Spending
* Financing

## Economic Trends

**Value vs. Price**

While most restaurant consumers are price-conscious, they are not only focused on consumer cost. Restaurants that are perceived as providing a higher-value offering are more likely to succeed than those that are just inexpensive. Customers have demonstrated a willingness to pay a little bit more for a meal that is significantly higher quality or has health benefits.

**Quick-Casual Growth**

The quick casual sector of restaurants is positioned to continue growing. These restaurants are slotted between fast food and casual dining restaurants. Many offers high-quality food focused in a narrow niche, and pricing that also falls between the two other types of restaurants. With more ethnic foods becoming mainstream and high-end chefs focusing on the sector, the sector has even more upward pressure.

## Data

**Required data**

State

**Data Storage Engine**

Due to the data nature it will be stored in relational database. Primary engine is MS SQL, though testing and demonstration will be done with Progress.

**Data Update**

Data can be updated using API calls, or upload of new data files.

Existing data will be updated, new data inserted (UPSERT).

Data load comment:

* Provided ETL is reentrant, previously loaded data is preserved, new data is inserted; for this data source PK is used
* Surrogate keys are used as PK in data storage, unique indexes are used to assure data integrity

## Deliverables

**Data Storage**

* Database with preloaded data

**Programs**

* Database Views
* REST API

**Data Sources**

Yelp Business Dataset, from Kaggle

<https://www.kaggle.com/yelp-dataset/yelp-dataset#yelp_academic_dataset_review.json>

Yelp API

<https://blog.rapidapi.com/yelp-fusion-api-profile-pull-local-business-data/>

Calling API

<https://api.yelp.com/v3/businesses/search?limit=2&location=Roma&locale=it_IT&categories=indpak&term=restaurant>

<https://blog.rapidapi.com/we-love-these-restaurant-apis/>

Consumer expenditure survey

<https://www.bls.gov/cex/>

BEA Personal Consumer Spending and Income by State (2019) 50 rows x

<https://www.bea.gov/data/income-saving/personal-income-by-state>

<https://www.bea.gov/data/consumer-spending/state>

### Data Usage Examples

**Use Case 1**

Customer is looking to open 10 new locations in the United States. They would like to start within one state and want to find best state based on consumer spending patterns and market saturation.

To support such criteria customer defines null hypothesis:

There is no difference between the 50 states in market saturation, ratings (average and spread) and consumer spending.

See online demo: <http://proximityone.com/pce.htm>

## Database Design

Restaurant

State

StatePCE

Restaurant  
Category

Category

**StatePCE**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Data Type** | **Key, Index** | **Comment** |
| **Id** | **INT NOT NULL** | **PK** | **Surrogate Key** |
| **StateId** | **INT NOT NULL** | **FK** |  |
| **Year** | **DATE NOT NULL** |  | **Year** |
| **OffPremisesFoodBeverages** | **INT NOT NULL** |  | **Sum $** |
| **ChangePct** | **DECIMAL (5,2) NOT NULL** |  |  |
| **EnteredDate** | **DATE NOT NULL** |  |  |
| **Source** | **VARCHAR (100)** |  | **Data Source** |

**State**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Data Type** | **Key, Index** | **Comment** |
| **Id** | **INT NOT NULL** | **PK** | **ANSI Code** |
| **Name** | **VARCHAR(50) NOT NULL** |  |  |
| **Name\_A2** | **CHAR(2) NOT NULL** |  | **Year** |
|  |  |  |  |
|  |  |  |  |

**Restaurant**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Data Type** | **Key, Index** | **Comment** |
| **Id** | **INT NOT NULL** | **PK** | **Surrogate key** |
| **BusinessId** | **VARCHAR(100) NOT NULL** | **UIDX** |  |
| **Name** | **VARCHAR(250) NOT NULL** |  |  |
| **StateId** | **INT NOT NULL** | **FK** |  |
| **City** | **VARCHAR(250) NOT NULL** |  |  |
| **Street** | **VARCHAR(250) NOT NULL** |  |  |
| **ZIP** | **VARCHAR(25) NOT NULL** |  |  |
| **Rating** | **DECIMAL(3,1) NOT NULL** |  | **Current value** |
| **CategoryId** | **INT NOT NULL** | **FK** |  |
| **Source** | **INT NOT NULL** |  | **Data Source \*1)** |
|  |  |  |  |
|  |  |  |  |

\*1) Use 1 for Yelp

**RestaurantCategory**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Data Type** | **Key, Index** | **Comment** |
| **Id** | **INT NOT NULL** | **PK** | **Surrogate Key** |
| **RestaurantId** | **INT NOT NULL** | **FK, UIDX1** |  |
| **CategoryId** | **INT NOT NULL** | **FK, UIDX1** |  |

**RestaurantCategory**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Data Type** | **Key, Index** | **Comment** |
| **Id** | **INT NOT NULL** | **PK** | **Surrogate Key** |
| **Name** | **VARCHAR(100) NOT NULL** | **UIDX** |  |
| **Description** | **VARCHAR(250)** |  |  |

**Comment to reader implementation:**

* Input – file name, sheet number or name, row with column name, column name
* Expects – state name in first column
* Find column id for expenditure
* Read all values into hash table

TODO

1. Write project document
2. Define data set, define data model
3. Define data sources
4. Pull data from data sources, API, etc.
5. Create database
6. Load data
7. Create data interface – View, Store Procedure, API