**BIA Final Project Plan – Jack Weston**

**Restaurant Analysis:** What restaurants are popular? What restaurants generate the highest revenue? Why?…

Questions / Brainstorming:

What kind of restaurants are the busiest and what metric are they based on?

* Expensive, cheap, moderate price? - price
* Location? - City
  + Median/Average income in the area will depend on success of restaurant? – Monthly income - Done
  + Occupation, dependents/family? - Occupation
* Which type of food is the most popular – cuisine - Done
* Seasonal Trends – order\_date Done
* Age range? – age - Done
* Do higher ratings lead to higher revenue? – rating - Done
* Does highest revenue mean highest quantity sold? – sales\_amount + sales\_qty - Done

**Hypothesis**

My hypothesis is higher income individuals with families will be the biggest spenders and the restaurants in big cities in neighborhoods with higher income individuals will have the highest revenue. The characteristics of some of these restaurants will include medium expensive items because customers are likely to come back sooner than at a high-end restaurant that may be an occasional event.

**Analysis**

Using the columns below I will plot graphs showing which restaurants are most popular by area, season, cuisine, location, customer demographic (Income, family size, marital status, etc.). I will use bar graphs, line graphs, tree maps etc. that deem fit for the best way to convey results.

Looking at the datasets I think there is room to use all 5 tables for this analysis and join tables together using identifiers:

1. menu\_id

1. f\_id
2. r\_id
3. user\_id
4. id

Columns that I will be looking at for analysis will include:

menu file:

1. cuisine - find which cuisine is popular
2. price - Is there a threshold on a cuisine and how popular it will be based on price

orders File:

1. order\_date - Are there seasonal trends to the data
2. sales\_amount - How much of a product are they selling
3. sales-qty - Same as above

restaurant file:

1. city - best location for a restaurant
2. rating – higher customer satisfaction leads to higher revenue

users file:

1. age - What age range goes to certain restaurants
2. Marital status - Do couples go out more?
3. Occupation - student vs full time employee
4. Monthly income - Assumption of higher income equals more money spent
5. Family size - bigger family equals bigger spenders

Data Cleaning:

1. Combined Chinese,North Indian with North Indian,Chinese
2. Combined Chinese,Indian with Indian,Chinese
3. Combined Biryani,North Indian and North Indian,Biryani
4. Combined North Indian,South Indian and South Indian,North Indian
5. Combined Biryani,Indian with Indian,Biryani
6. Combined North Indian,Indian with Indian,North Indian
7. Combine South Indian,Chinese with Chinese,South Indian
8. Combine Chinese,Biryani with Biryani,Chinese