

Report on Subway | Python Code, Outputs & Analysis

Written By Adri Mukherjee

adri.mukherjee1@gmail.com

"Please do not hesitate to reach out if you have questions or want to chat about other approaches to Subway's Analysis"

Table of Contents



1. [Analysis 1: Identifying the Issue - Are the Subway Ratings Improving?](#)
2. [Analysis 2: Competitive Analysis – Is This Performance Usual for Competitors, Or Is It Only A Subway Issue?](#)
3. [Analysis 3: Analyzing Market Position - Do National Chains Usually Face Lower Ratings Compared to Small/Local/Boutique Restaurants?](#)
4. [Analysis 4: Testing the Reliability of Reviews – Do Customers Only Post Reviews When They Are Either Very Happy or Very Angry with The Service, Not In Between?](#)
5. [Analysis 5: How Are the Ratings like For Local Businesses VS. National Chains?](#)
6. [Analysis 6: Comparing Average Ratings for Subway VS. Competitors Over Time](#)
7. [Overall Findings – Reporting to CEO](#)

##Analysis 1: Identifying the Issue - Are the Subway Ratings Improving?

Take year from review date

```
reviews_data['year'] = pd.to_datetime(reviews_data['date']).dt.year
```

Group by year -> avg rating & count of reviews

```
yearly_summary = reviews_data.groupby('year').agg(  
    avg_rating=('stars', 'mean'),  
    num_ratings=('stars', 'count')  
) .reset_index()
```

```
import matplotlib.pyplot as plt
```

```
fig, ax1 = plt.subplots(figsize=(10, 6)) #Plotting & adjusting scale
```

Primary y-axis: Average rating over time

```
ax1.plot(yearly_summary['year'], yearly_summary['avg_rating'], color='blue', label='Average Rating',  
marker='o')
```

```
ax1.set_xlabel('Year')
```

```
ax1.set_ylabel('Average Rating', color='blue')
```

```
ax1.tick_params(axis='y', labelcolor='blue')
```

```
ax1.set_title('Average Rating and Number of Ratings Over Time')
```

Secondary y-axis: Number of ratings

```
ax2 = ax1.twinx()
```

```
ax2.bar(yearly_summary['year'], yearly_summary['num_ratings'], color='gray', alpha=0.5, label='Number  
of Ratings')
```

```
ax2.set_ylabel('Number of Ratings', color='gray')
```

```
ax2.tick_params(axis='y', labelcolor='gray')
```

```
# Legends and grid
```

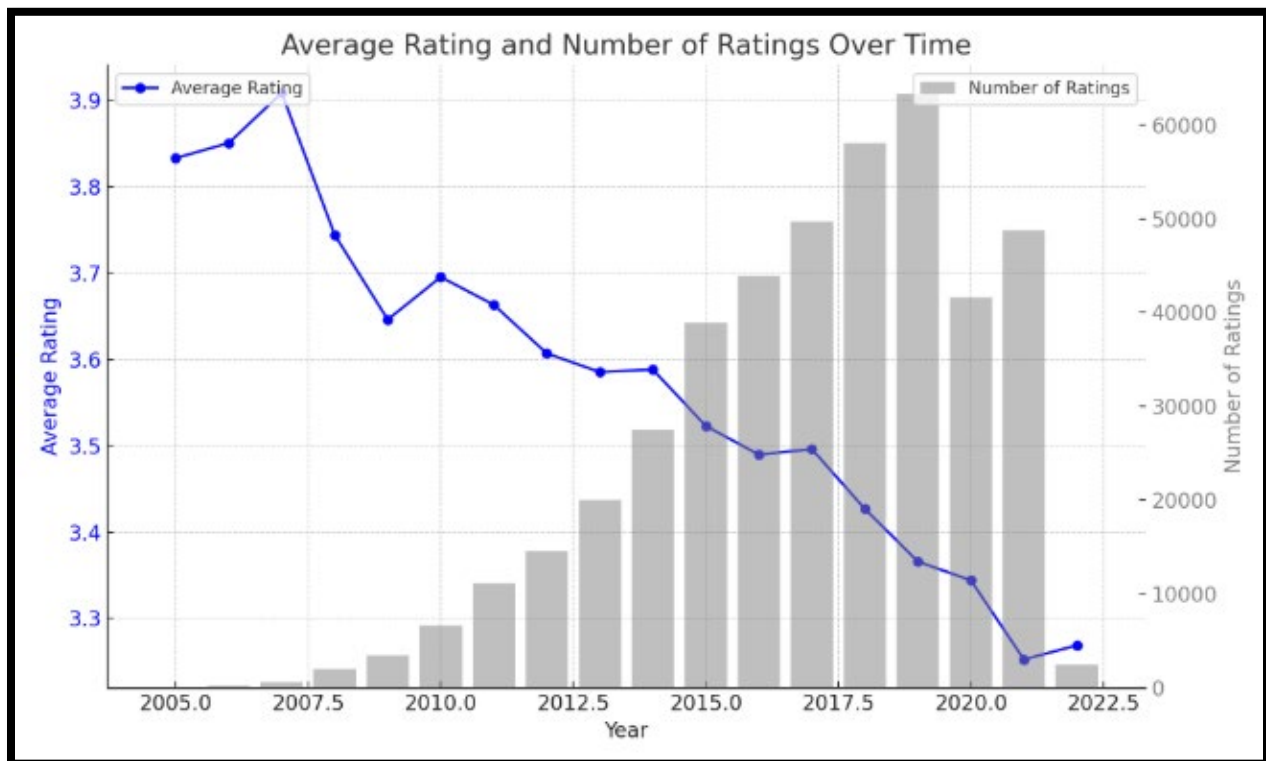
```
fig.tight_layout()
```

```
ax1.legend(loc='upper left')
```

```
ax2.legend(loc='upper right')
```

```
plt.grid()
```

```
plt.show()
```



~ Summarized Findings ~

As shown in the graph, we see a negative trend in average ratings over the years, indicating no improvement in ratings.

##Analysis 2: Competitive Analysis – Is This Performance Usual for Competitors, Or Is It Only A Subway Issue?

Subway and competitors using categories

```
subway_data = restaurants_data[restaurants_data['name'].str.contains('Subway', case=False)]
```

```
competitor_names = ['Jimmy John', 'Jersey Mike'] # Competitors
```

```
competitor_data = restaurants_data[restaurants_data['name'].str.contains('|'.join(competitor_names), case=False)]
```

Review filtering

```
subway_reviews =  
reviews_with_state[reviews_with_state['business_id'].isin(subway_data['business_id'])]
```

```
competitor_reviews =  
reviews_with_state[reviews_with_state['business_id'].isin(competitor_data['business_id'])]
```

Mean and std dev for ratings

```
comparison_stats = pd.concat([  
    subway_reviews.assign(brand='Subway'),  
    competitor_reviews.assign(brand='Competitor')  
]).groupby('brand').agg(  
    mean_rating=('stars', 'mean'),  
    std_rating=('stars', 'std')  
) .reset_index()
```

Plotting

```
import matplotlib.pyplot as plt
```

```
fig, ax = plt.subplots(figsize=(8, 6))
```

Bar plot for mean

```
ax.bar(comparison_stats['brand'], comparison_stats['mean_rating'], color='blue', alpha=0.7,  
label='Mean Rating')
```

```
# Error bars for standard deviation
```

```
ax.errorbar(comparison_stats['brand'], comparison_stats['mean_rating'],  
            yerr=comparison_stats['std_rating'], fmt='o', color='black', label='Standard Deviation')
```

```
#Labels
```

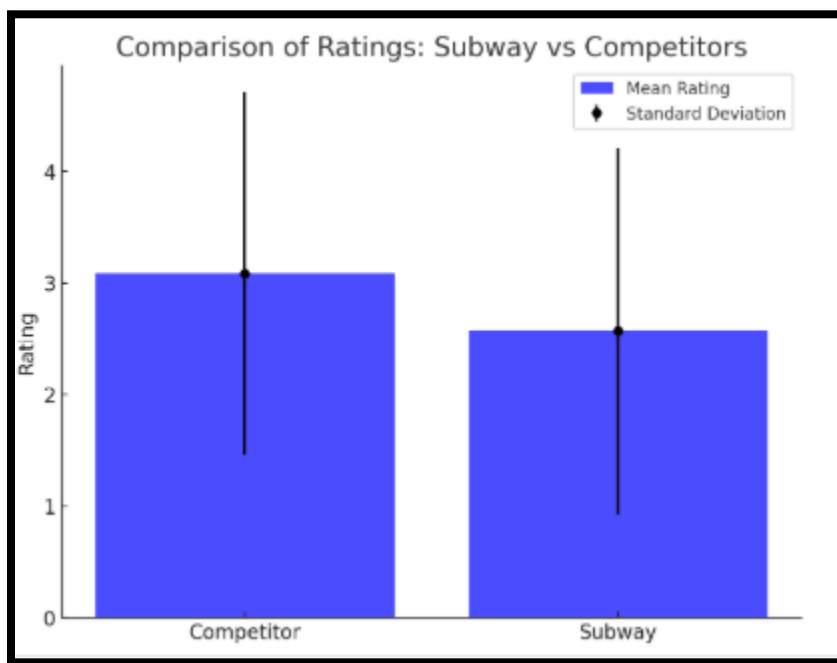
```
ax.set_ylabel('Rating')
```

```
ax.set_title('Comparison of Ratings: Subway vs Competitors')
```

```
ax.legend()
```

```
plt.grid()
```

```
plt.show()
```



~ Summarized Findings ~

The standard deviation in ratings is similar between Subway and its competitors, but Subway has a lower mean rating (2.57 vs 3.09). Poor ratings might not necessarily mean there is an issue with sandwiches, but it is specific to Subway.

##Analysis 3: Analyzing Market Position - Do National Chains Usually Face Lower Ratings Compared to Small/Local/Boutique Restaurants?

Restaurant presence in cities

```
city_presence = restaurants_data.groupby('name')['city'].nunique().reset_index()
city_presence.rename(columns={'city': 'city_count'}, inplace=True)
```

Conditions to determine national, local, and regional

```
restaurants_data = restaurants_data.merge(city_presence, on='name', how='left')
restaurants_data['chain_category'] = restaurants_data['city_count'].apply(
    lambda x: 'National Chain' if x > 50 else 'Local Chain' if x == 1 else 'Regional Chain'
)
```

Add chain data into reviews

```
reviews_with_chains = reviews_with_state.merge(
    restaurants_data[['business_id', 'chain_category']], on='business_id', how='left'
)
```

Comparing averages

```
chain_comparison = reviews_with_chains.groupby('chain_category').agg(
    avg_rating=('stars', 'mean'),
    std_rating=('stars', 'std')
).reset_index()
```

Plot – Does average rating decrease as restaurant size increases?

```
import matplotlib.pyplot as plt
fig, ax = plt.subplots(figsize=(8, 6))
```

```
# Bar plot
```

```
ax.bar(chain_comparison['chain_category'], chain_comparison['avg_rating'], color='blue', alpha=0.7,  
label='Mean Rating')
```

```
# Error bars
```

```
ax.errorbar(chain_comparison['chain_category'], chain_comparison['avg_rating'],  
            yerr=chain_comparison['std_rating'], fmt='o', color='black', label='Standard Deviation')
```

```
#Labels
```

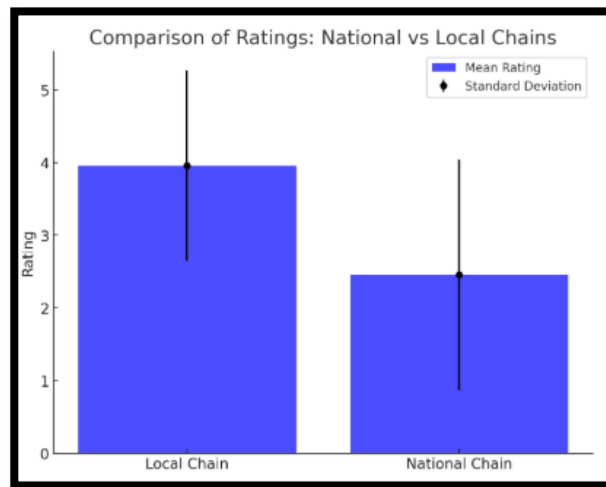
```
ax.set_ylabel('Rating')
```

```
ax.set_title('Comparison of Ratings: National vs Local Chains')
```

```
ax.legend()
```

```
plt.grid()
```

```
plt.show()
```



~ Summarized Findings ~

Local chains have average rating of 3.96 and national chains have 2.46. As a result, local chains tend to get higher ratings on average. National chains have higher standard deviation, which shows the wider range of user reviews compared to local ones. Therefore, the data supports the claim that smaller restaurants get higher ratings. Average rating decreases as the restaurant size increases. In order from highest average ratings to lowest, it goes from local chains, smaller regional chains, and then national chains.

##Analysis 4: Testing the Reliability of Reviews – Do Customers Only Post Reviews When They Are Either Very Happy or Very Angry with The Service, Not In Between?

```
# Count rating freq
```

```
rating_distribution = reviews_data['stars'].value_counts().sort_index().reset_index()
```

```
rating_distribution.columns = ['rating', 'count']
```

```
# Plotting
```

```
import matplotlib.pyplot as plt
```

```
fig, ax = plt.subplots(figsize=(10, 6))
```

```
# Bar plot
```

```
ax.bar(rating_distribution['rating'], rating_distribution['count'], color='blue', alpha=0.7)
```

```
ax.set_xlabel('Ratings')
```

```
ax.set_ylabel('Number of Reviews')
```

```
ax.set_title('Distribution of Ratings')
```

```
plt.grid()
```

```
plt.show()
```

```
# Filter for any year range
```

```
# Filter data for 2018-2021
```

```
filtered_data = reviews_data[reviews_data['year'].between(2018, 2021)]
```

```
# Count ratings freq per year
```

```
yearly_rating_distribution = filtered_data.groupby(['year', 'stars']).size().unstack(fill_value=0)
```



```
# Normal dist
```

```
yearly_rating_percentage = yearly_rating_distribution.div(yearly_rating_distribution.sum(axis=1),  
axis=0) * 100
```

```
fig, ax = plt.subplots(figsize=(12, 8))
```

```
yearly_rating_percentage.plot(kind='bar', stacked=True, ax=ax, colormap='viridis', alpha=0.8) #Color  
coded for each rating
```

```
ax.set_xlabel('Year')
```

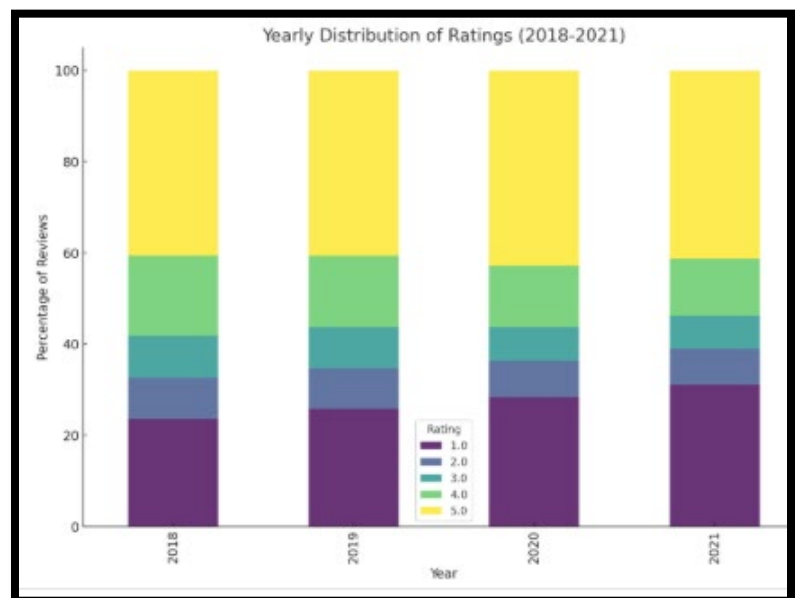
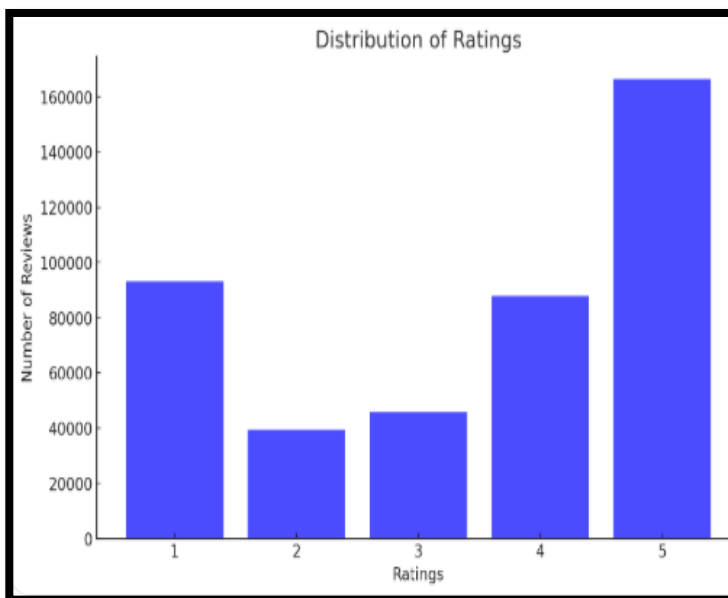
```
ax.set_ylabel('Percentage of Reviews')
```

```
ax.set_title('Yearly Distribution of Ratings (2018-2021)')
```

```
plt.legend(title='Rating')
```

```
plt.grid()
```

```
plt.show()
```



~ Summarized Findings ~

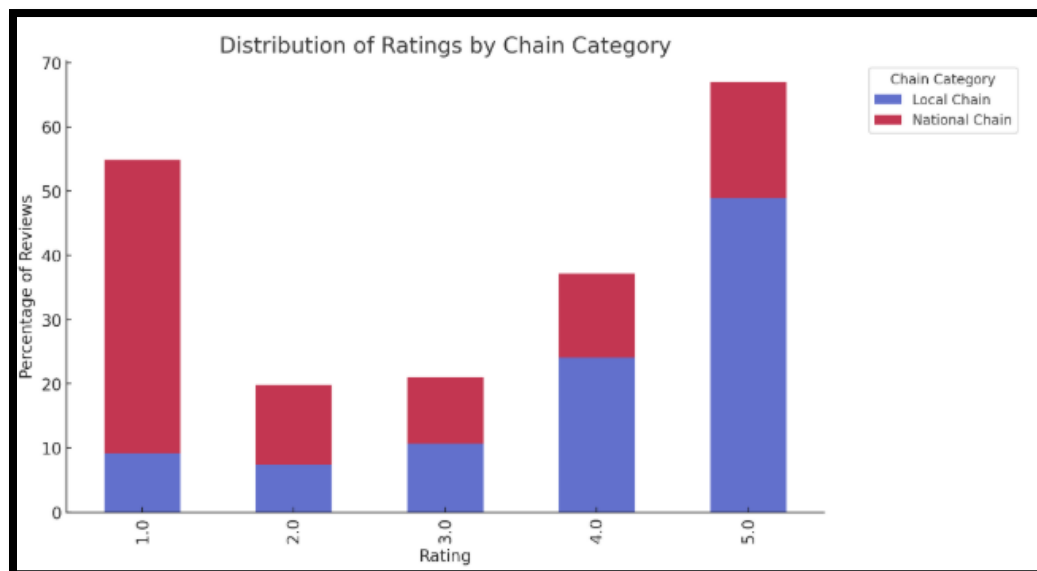
Data slightly supports the statement. The bar plot shows how skewed the data is towards the extreme ratings (1 or 5 stars). This could either mean that customers tend to report only the extreme ends of their experiences, or that the service is mainly 1- or 5-star worthy.

##Analysis 5: How Are the Ratings like For Local Businesses VS. National Chains?

```
chain_rating_distribution = reviews_with_chains.groupby(['chain_category',  
'stars']).size().unstack(fill_value=0)  
  
chain_rating_percentage = chain_rating_distribution.div(chain_rating_distribution.sum(axis=1), axis=0)  
* 100
```

#Plot

```
fig1, ax1 = plt.subplots(figsize=(10, 6))  
  
chain_rating_percentage.T.plot(kind='bar', stacked=True, ax=ax1, colormap='coolwarm', alpha=0.8)  
  
ax1.set_xlabel('Rating')  
ax1.set_ylabel('Percentage of Reviews')  
ax1.set_title('Distribution of Ratings by Chain Category')  
plt.legend(title='Chain Category', bbox_to_anchor=(1.05, 1))  
plt.grid()
```



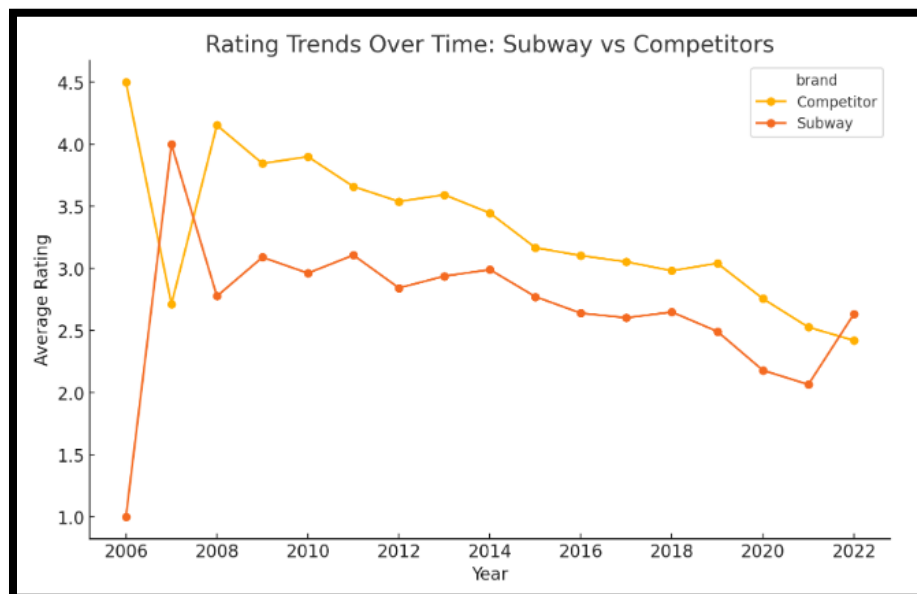
~ Summarized Findings ~

Like most national chains, Subway experiences extreme ratings (primarily 1-star) compared to local chains. This comes down to demographics and the target customers for each location of a national chain restaurant. For instance, service from a local small business would be much more prioritized than that of a national chain, which is where the room for improvement comes into play.

##Analysis 6: Comparing Average Ratings for Subway VS. Competitors Over Time

```
comparison_over_time = pd.concat([
    subway_reviews.assign(brand='Subway'),
    competitor_reviews.assign(brand='Competitor')
]).groupby(['year', 'brand'])['stars'].mean().unstack()

# Plot
fig2, ax2 = plt.subplots(figsize=(10, 6))
comparison_over_time.plot(ax=ax2, marker='o')
ax2.set_xlabel('Year')
ax2.set_ylabel('Average Rating')
ax2.set_title('Rating Trends Over Time: Subway vs Competitors')
plt.grid()
```



~ Summarized Findings ~

We see here that Subway has gotten worse reviews than its Competitors for every year except 2008. This shows that its nothing to do with the sandwich business being difficult, but rather an issue with Subway service.

Overall Findings – Reporting to CEO

John should worry since Subway's average rating of 2.57 is a lot lower than the competitive average of 3.09. A 4.5/5 target is unrealistic for a national chain, as we see local chains average 3.96 while national chains average 2.46. John should target improvements in service and customer experience and target a 3.5 score at first. The data shows that ratings are not consistently improving, and that it is skewed towards the 1- and 5-star ratings. Targeting underperforming stores and prioritizing quality could improve customer satisfaction.