**TASK 1: Top 10 Areas with Most Restaurants**

**Objective: Identify the top 10 areas with the highest number of restaurants.**

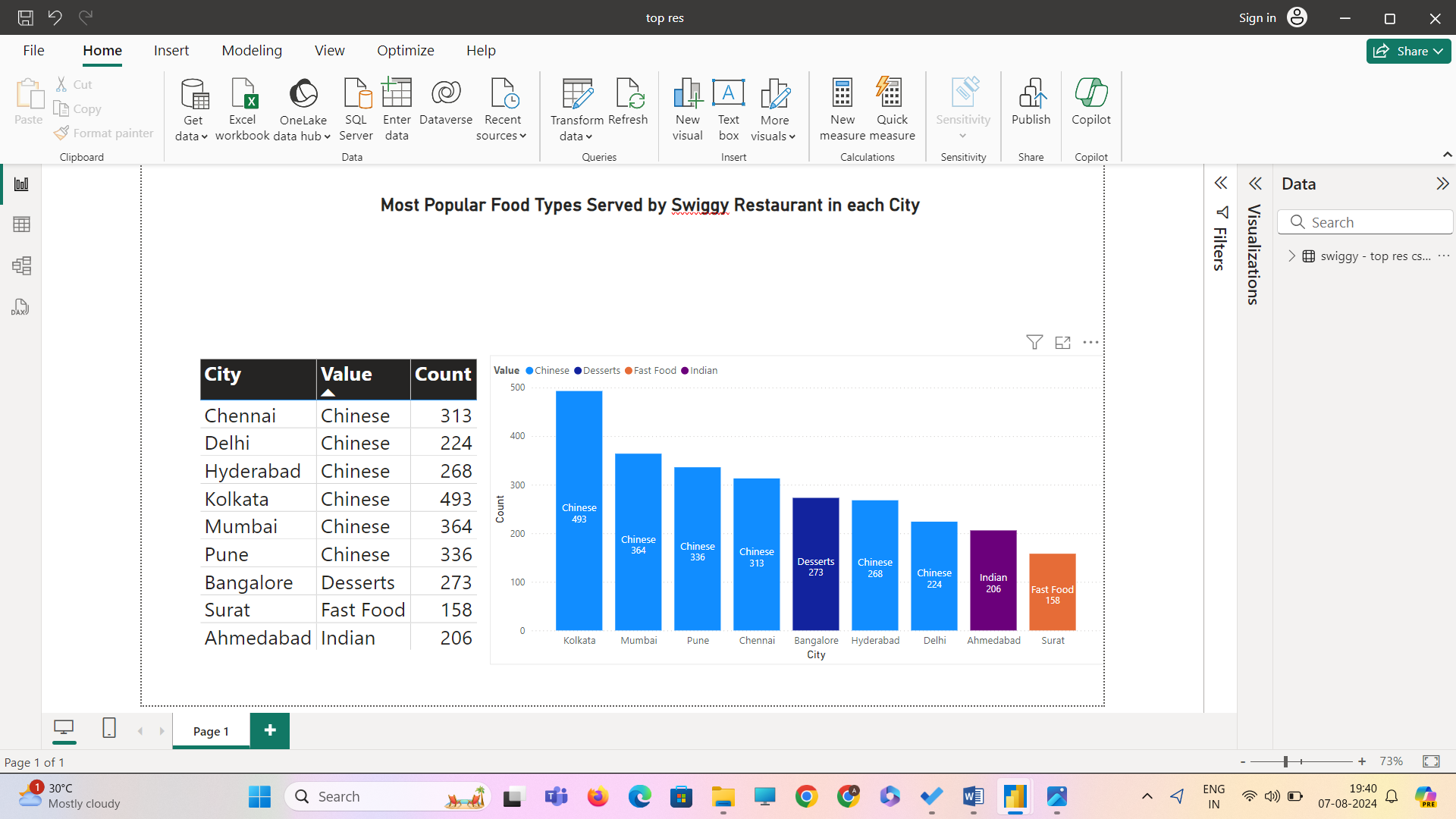
****

**Interpretation:**

* The area with highest number of restaurants: Rohini

**Task 2: Most Popular Food Types Served by Swiggy Restaurants in Each City**

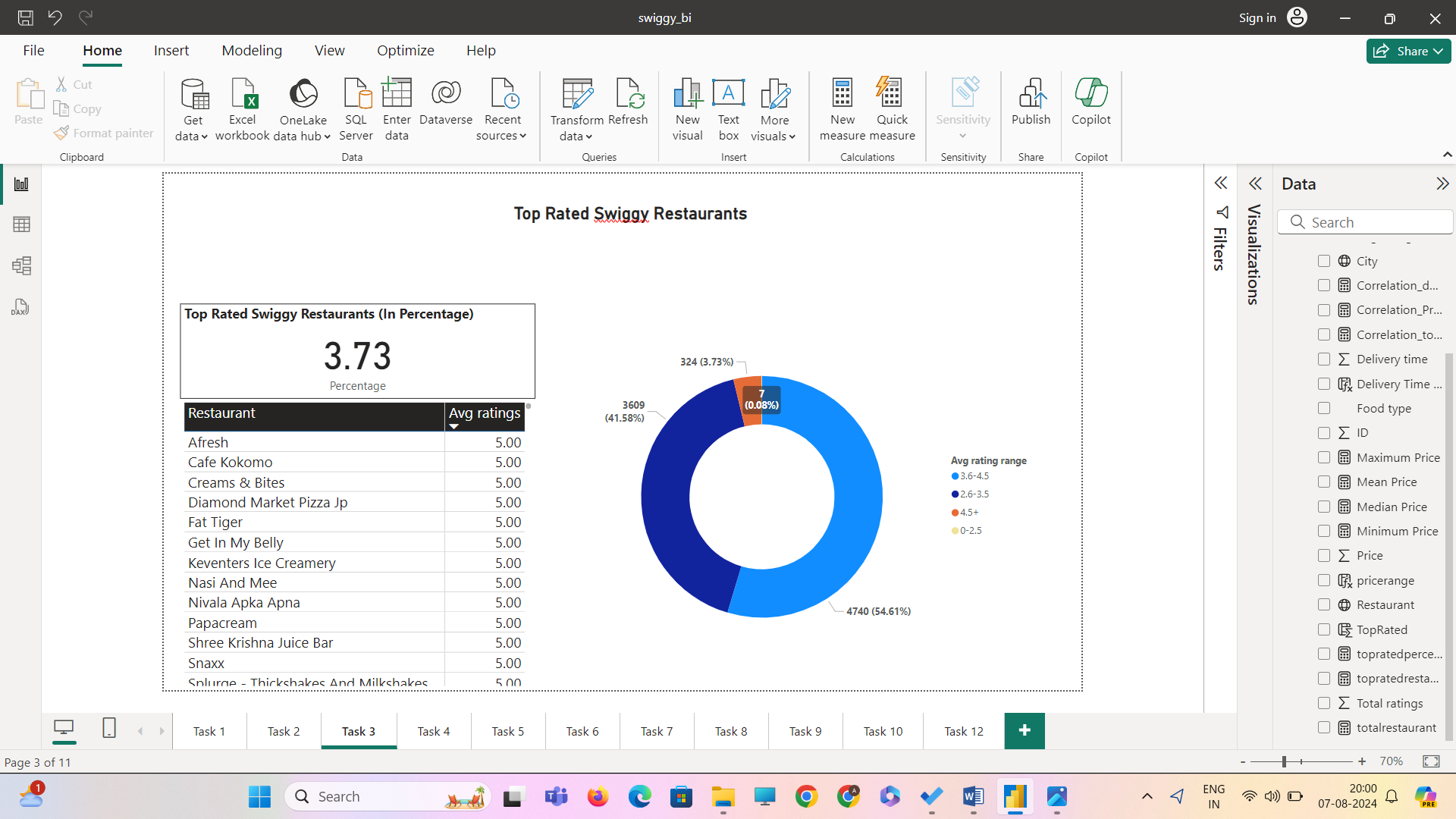
**Objective: Determine the most popular food types served in each city.**

****

**Interpretation:** In6 out of 9 cities **Chinese** is the highest food type served . Chinese as a food type in India is appropriately called Indo-chinese . It requires less ingredients to make Chinese food compared to Indian cuisine , and Chinese cuisine is cooked faster and served faster.

**Task 3: Top Rated Swiggy Restaurants (In Percentage)**

**Objective: Find the percentage of top-rated restaurants (e.g., those with an average rating above 4.5).**

****

New Column using DAX function

TopRated = IF('swiggy - swiggy csv csv'[Avg ratings] > 4.5, 1, 0)

New Measures created using DAX functions

totalrestaurant = countrows('swiggy - swiggy csv csv')

topratedrestaurants = sum('swiggy - swiggy csv csv'[TopRated])

topratedpercentage = DIVIDE([topratedrestaurants],[totalrestaurant],0)\*100

further , avg rating range is created as column using

DAX function

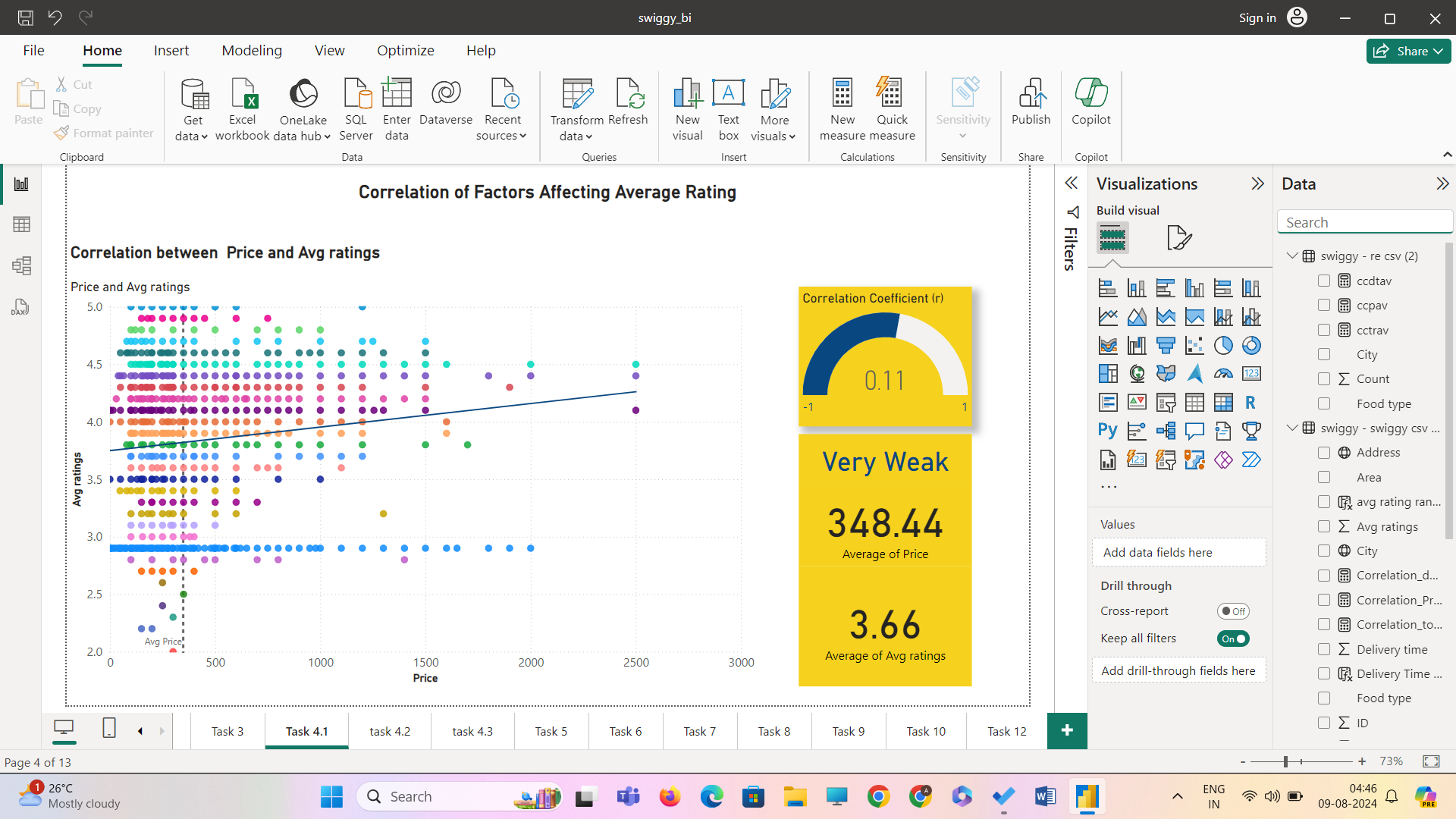
avg rating range = switch( true(),'swiggy - swiggy csv csv'[Avg ratings]<=2.5,"0-2.5",'swiggy - swiggy csv csv'[Avg ratings]<=3.5,"2.6-3.5",'swiggy - swiggy csv csv'[Avg ratings]<=4.5,"3.6-4.5",'swiggy - swiggy csv csv'[Avg ratings]>4.5,"4.5+")

To Create a DONUT chart and grouped with tables (Restaurant and Avg Rating) , which will be helpful in future analysis.

**Task 4: Correlation of Factors Affecting Average Rating**

**Objective: Identify correlations between different factors (e.g., price, total ratings, delivery time) and average rating**.

**Correlation between price and avg ratings**



New measure created using DAX function to find r

Correlation\_Price\_AvgRating =

VAR AvgX = AVERAGE('swiggy - swiggy csv csv'[Price])

VAR AvgY = AVERAGE('swiggy - swiggy csv csv'[Avg ratings])

VAR Num = SUMX('swiggy - swiggy csv csv', ('swiggy - swiggy csv csv'[Price] - AvgX) \* ('swiggy - swiggy csv csv'[Avg ratings] - AvgY))

VAR DenomX = SQRT(SUMX('swiggy - swiggy csv csv', ('swiggy - swiggy csv csv'[Price] - AvgX) ^ 2))

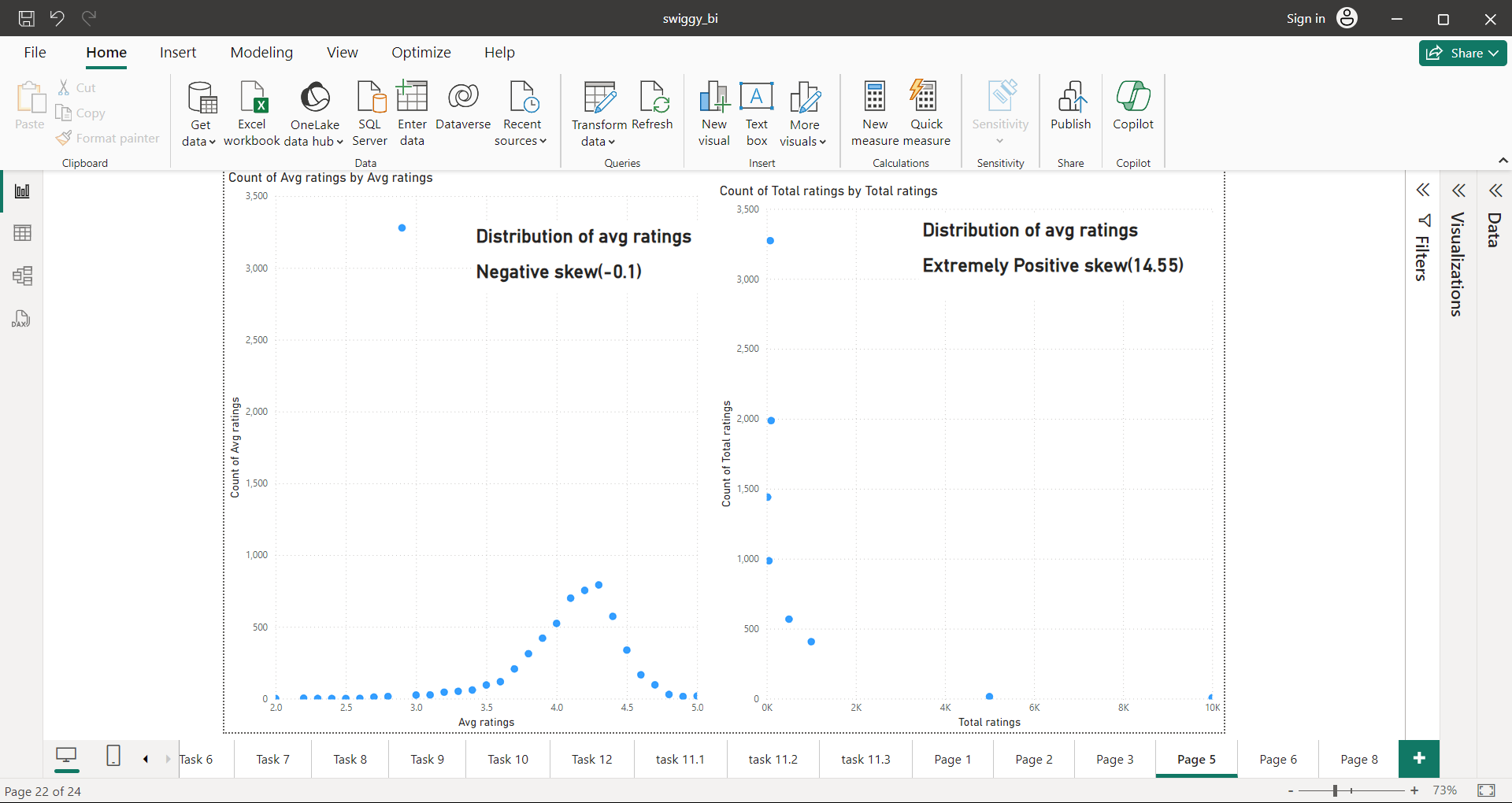
VAR DenomY = SQRT(SUMX('swiggy - swiggy csv csv', ('swiggy - swiggy csv csv'[Avg ratings] - AvgY) ^ 2))

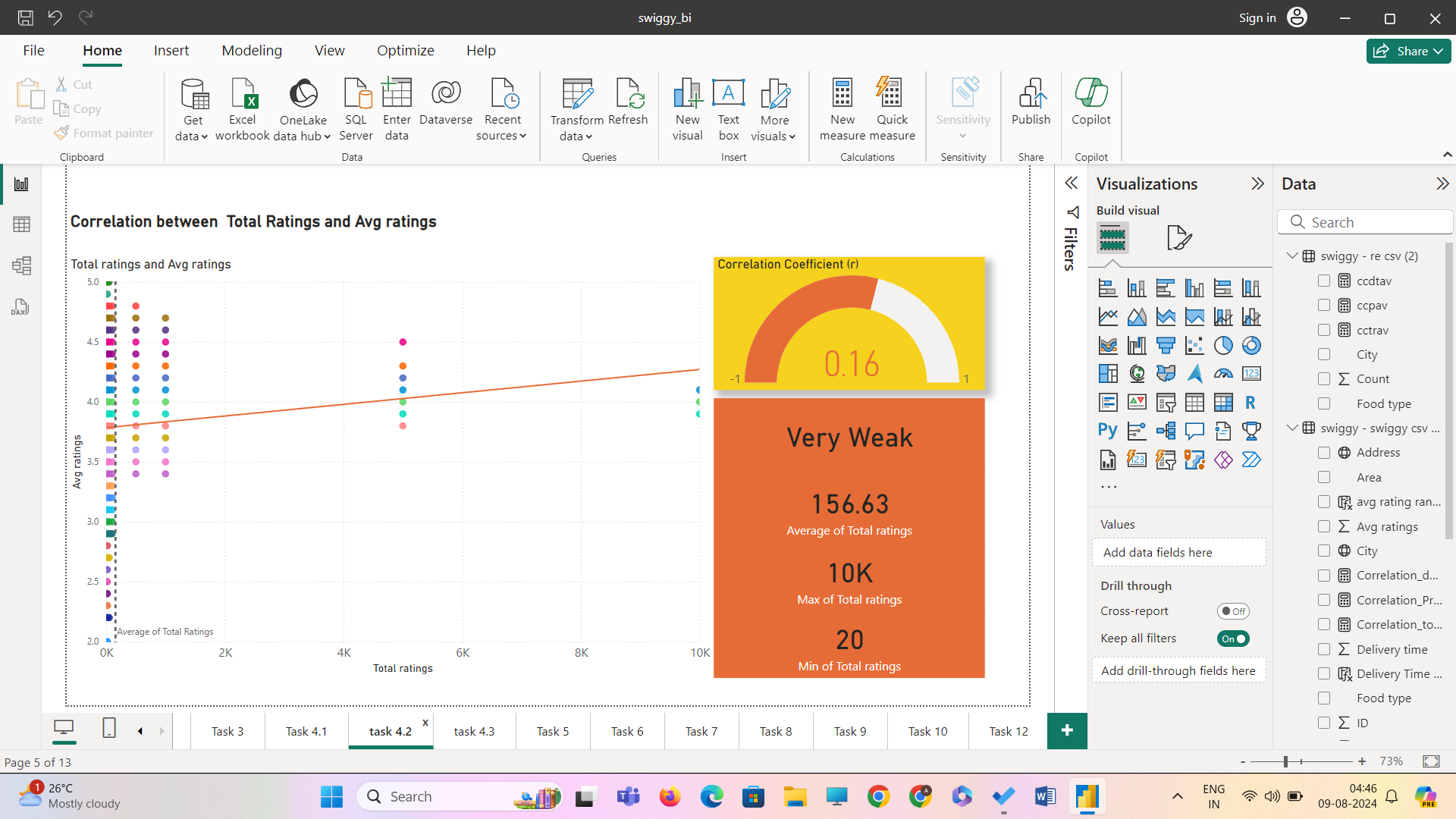
RETURN Num / (DenomX \* DenomY)

**Interpretation:** There is a very weak positive correlation between price and avg ratings of restaurants.

**NOTE:**both fields are skewed in opposite so the r value lacks to capture the nuances introduced by skewness and it measures only the linear relationship and not the shape of distribution.

There are points where price is low and the avg ratings are higher and there are points where the price is higher and the avg ratings are low . in the end , the customer ratings and feedback depends not only price but also depends on food taste , value of product , packaging , quality of service and overall satisfaction

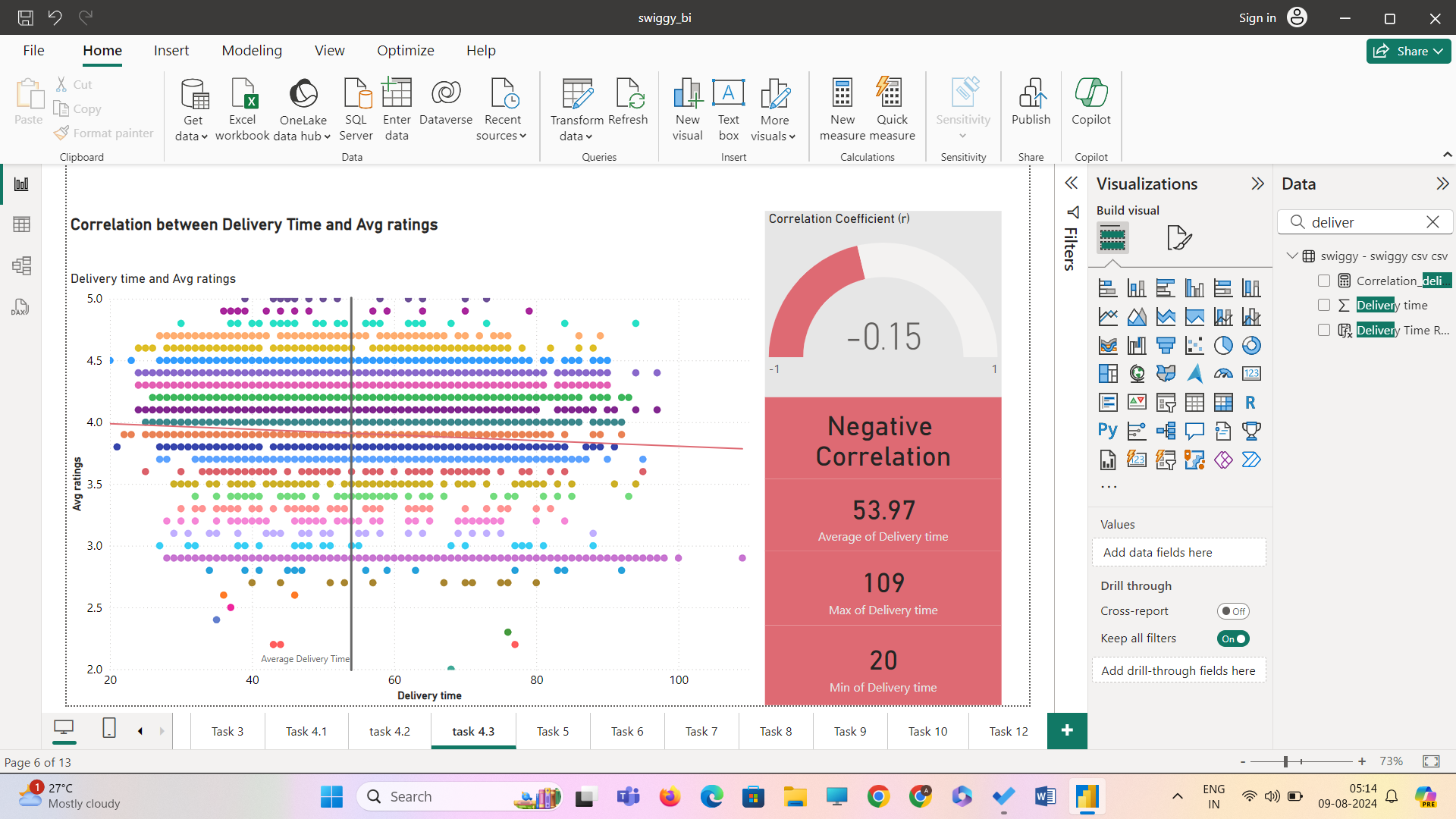


**Correlation between Total Ratings and Avg Ratings**

**Interpretation:** The Correlation between total ratings and avg ratings is very weak and positive. The skewness of the total rating is 14.55 which is an extremely right skewed distribution. It has extreme distribution . The relationship between total ratings and avg rating is very complex.

High total rating doesn't mean high avg ratings , it depends on individual customer ratings . whether a customer rates a business 1 star or 2 star or 5 star.in simple terms it is not only the quantity,but quality.

**Correlation between Delivery Time and Avg Ratings**



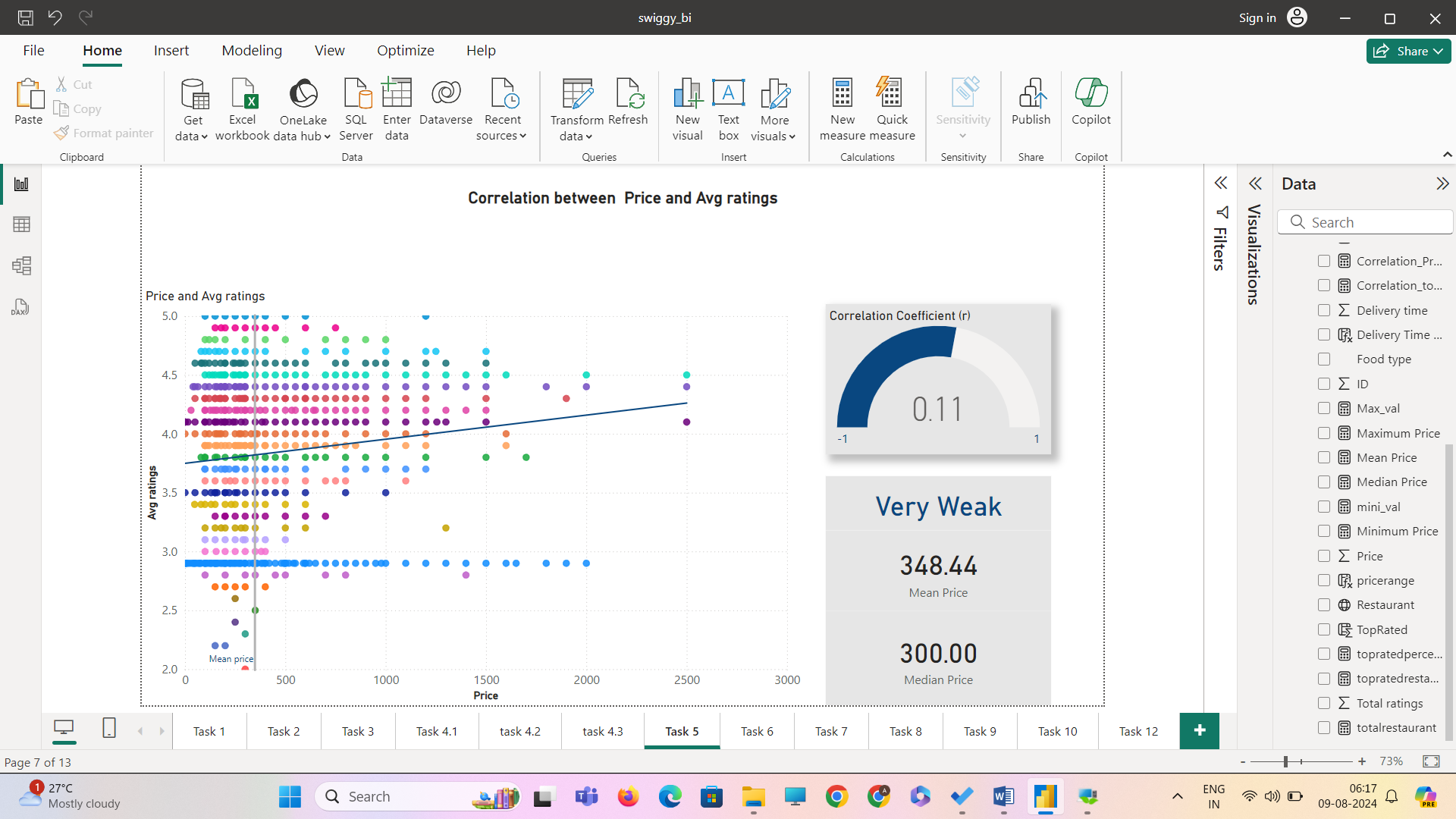
**Interpretation:**

The correlation between delivery time and avg ratings is weak negative correlation only measuring the linear relationship. The distribution of delivery time is slightly right skewed(0.378) forming almost a bell shape and the distribution of avg rating is left skewed so the relationship is determined as potential non-linearity.

There are points where delivery time is low and avg ratings are higher showing signs of negative correlation, which is normal and there are also points where delivery time is higher and avg ratings also higher showing signs of positive correlation . The reasoning behind high avg rating even in the case of high delivery time is long distance from restaurant to customer location , swiggy gives the approx estimated delivery time even before ordering from places to customer. The other reason is avg rating also influenced by quality of food , taste , packaging ,etc which is also the reason for low delivery time but low avg rating showing positive correlation.

**Task 5: Correlation Between Restaurant Price and Average Rating**

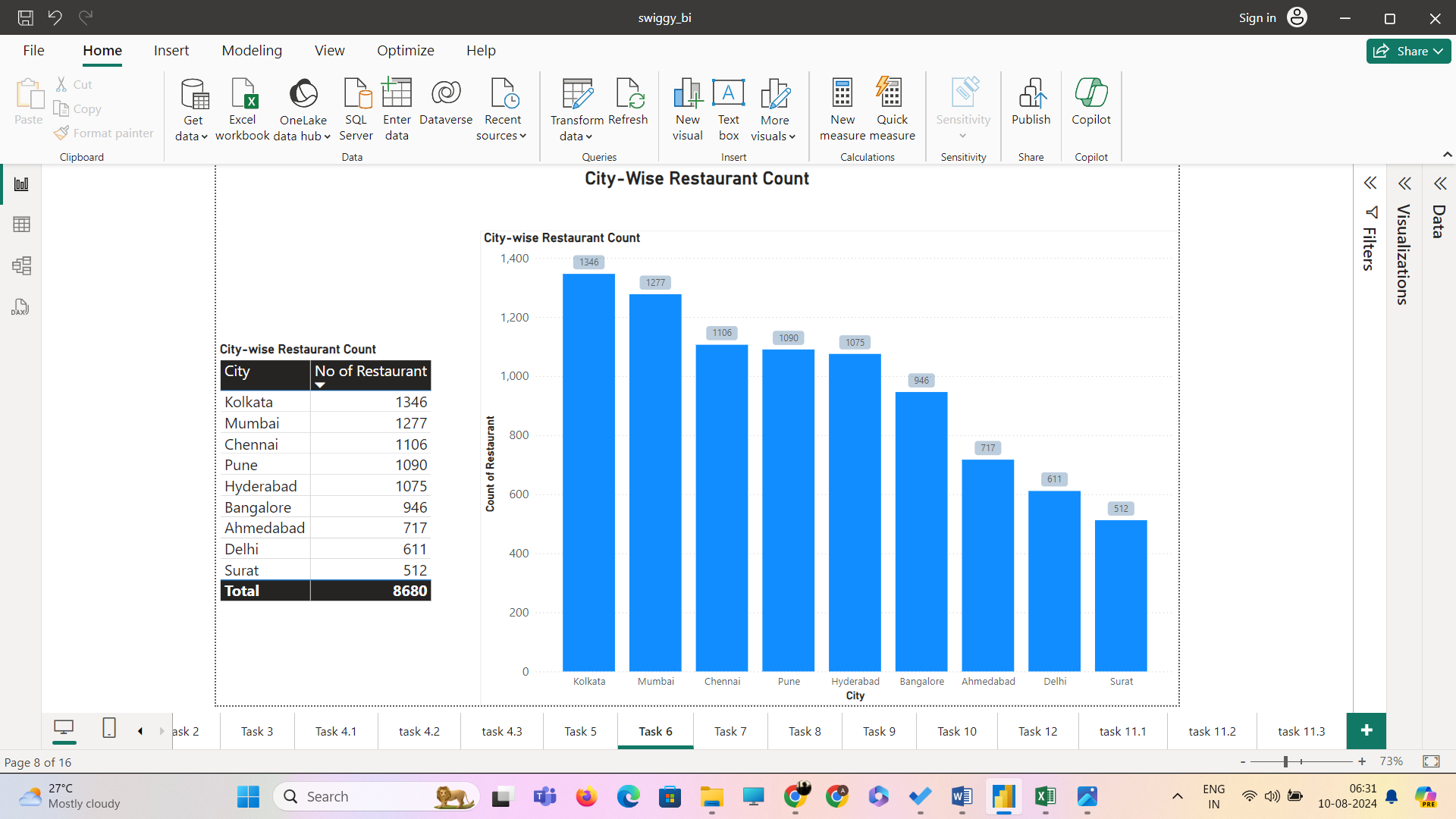
**Objective**: Analyse the relationship between restaurant price and average rating.

****

**Interpretation:**

* The correlation between price and avg ratings is very weak but positive .
* The mean and median of price suggest the distribution is positively skewed.
* The mean(3.66) and median (3.90)of avg rating suggest the distribution is negatively skewed.
* There are outliers in the both price and avg ratings creating pull and variability(high).
* The relationship cannot be defined by the r value , due to skewness in price and avg ratings ,it is a non-linear relationship.

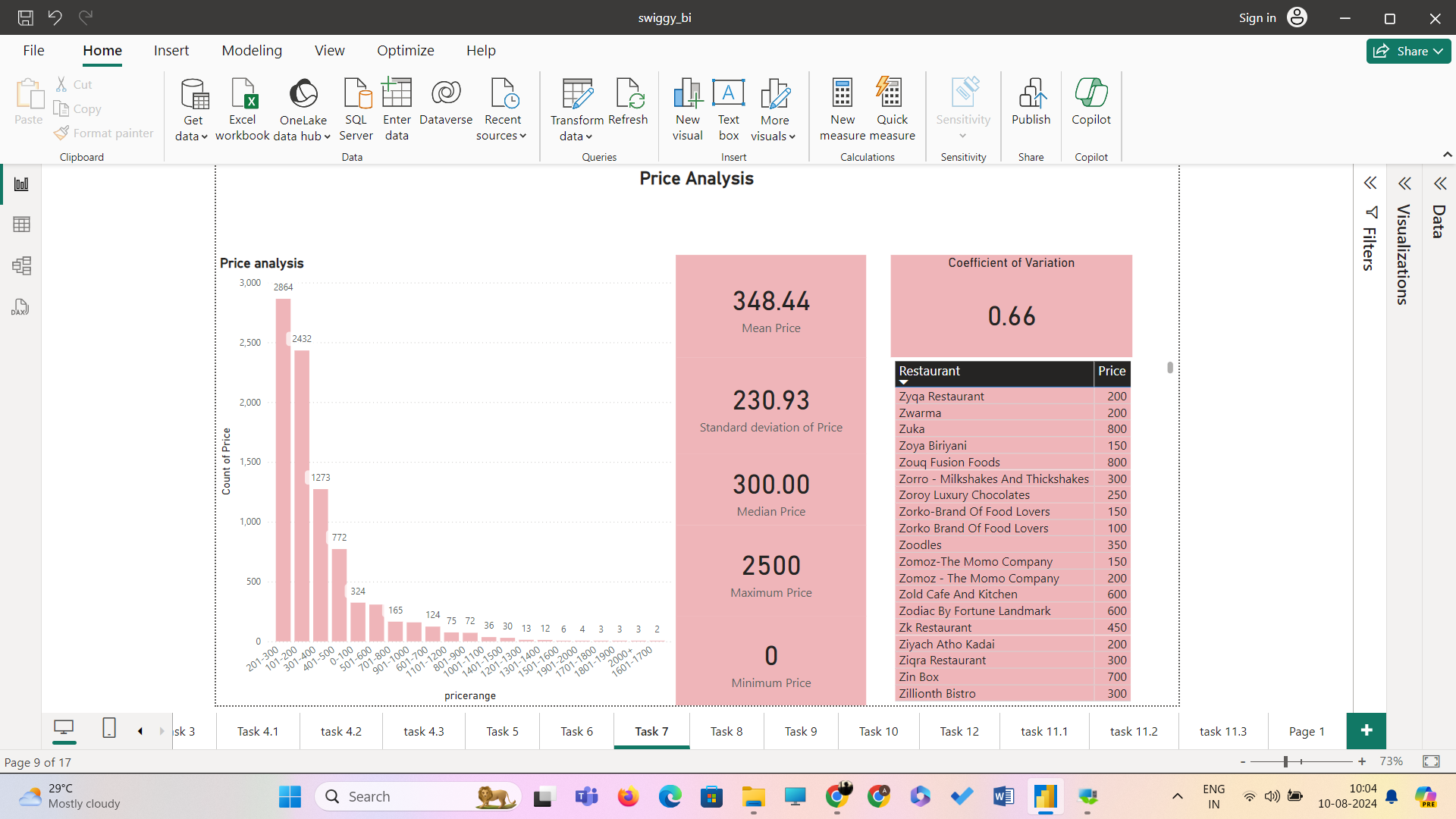
**Task 6: City-wise Restaurant Count**

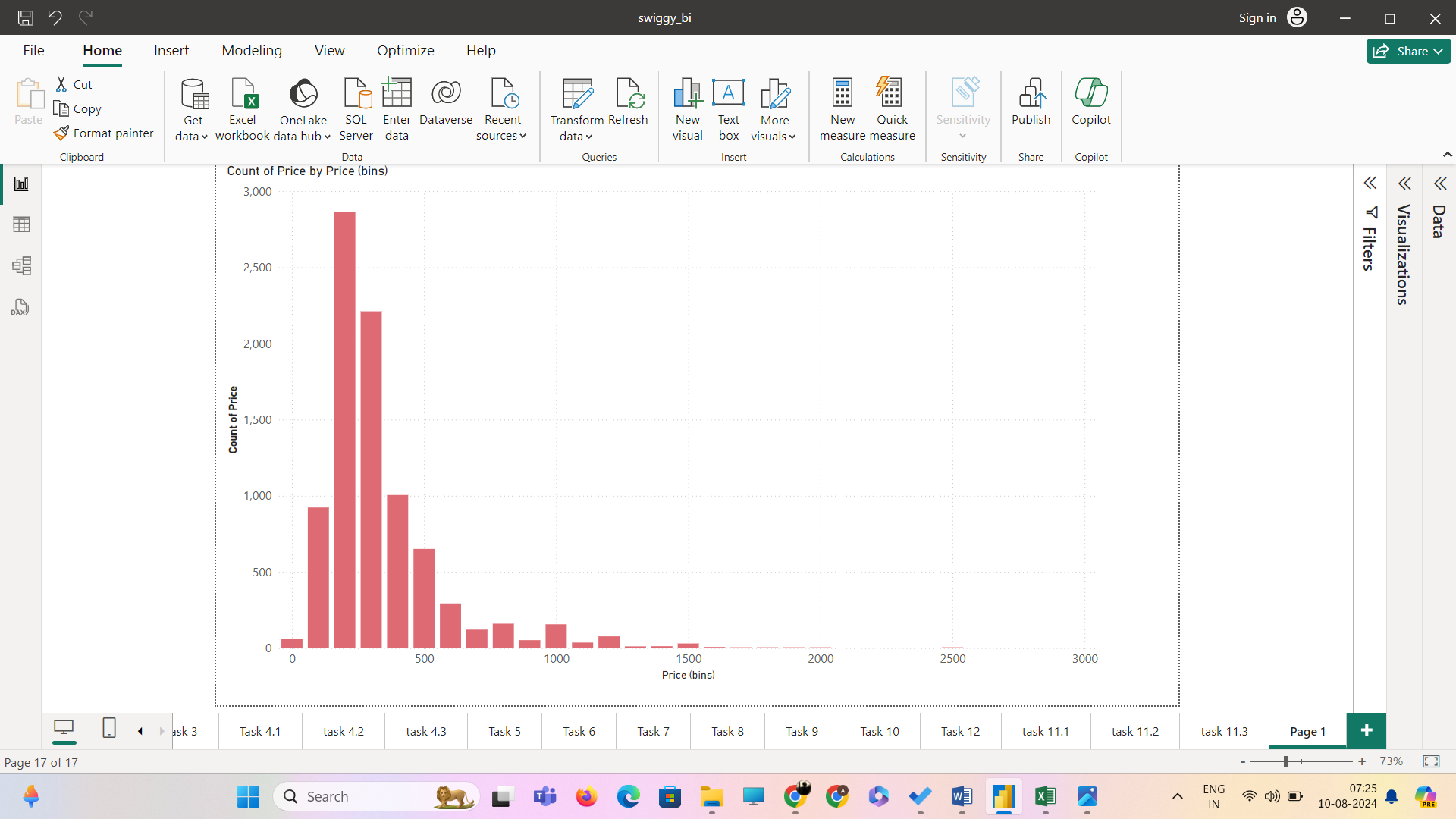
**Objective**: Find out the number of restaurants in each city.

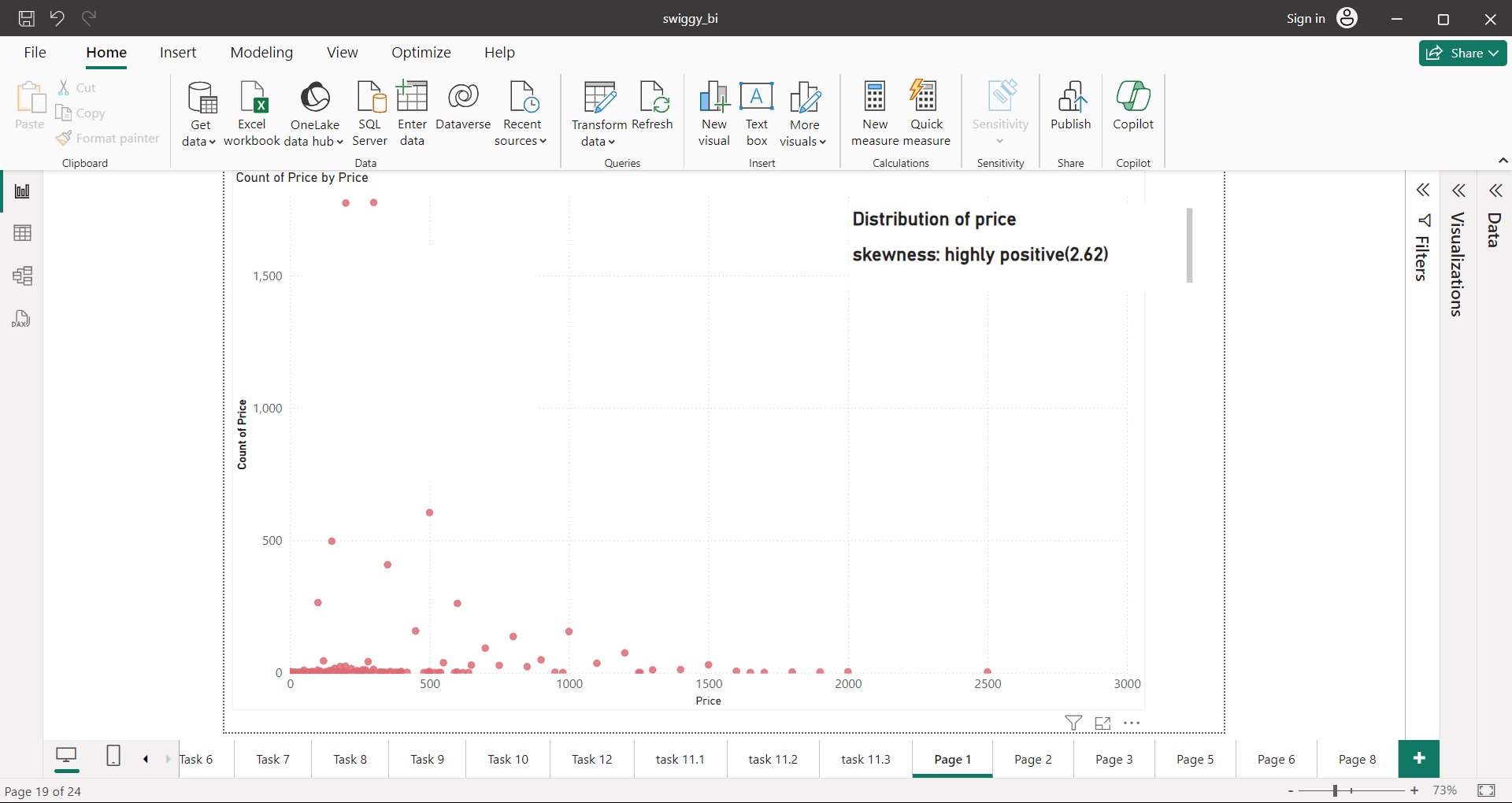
**Interpretation:**

* Kolkata is the city with the most number of restaurants.
* Surat is the city with the least number of restaurants.

**Task 7: Price Analysis**

**Objective**: Analyse the price distribution of restaurants.

Note: New Column named pricerange is created

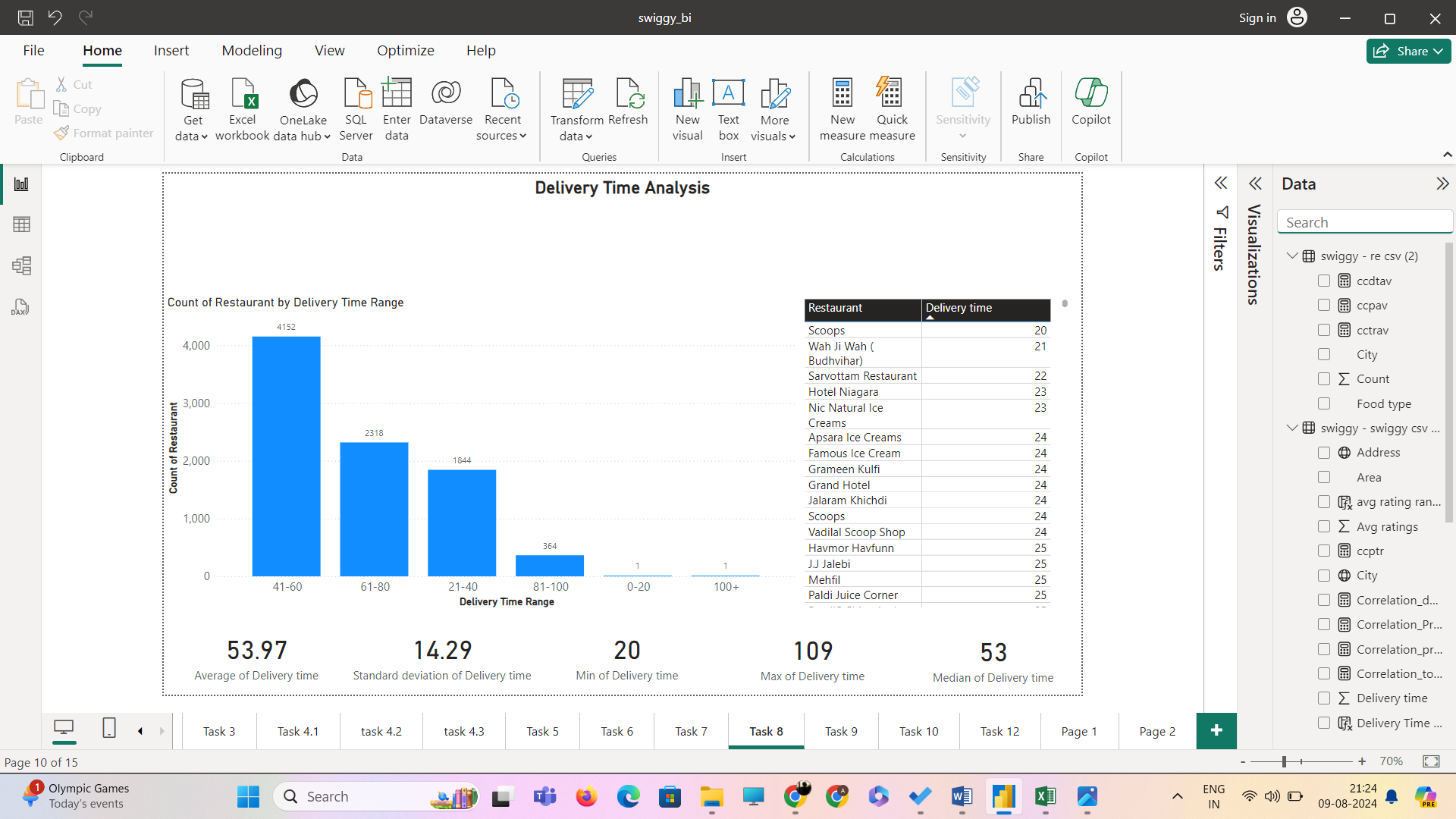
****

**Analysis and Interpretation :**

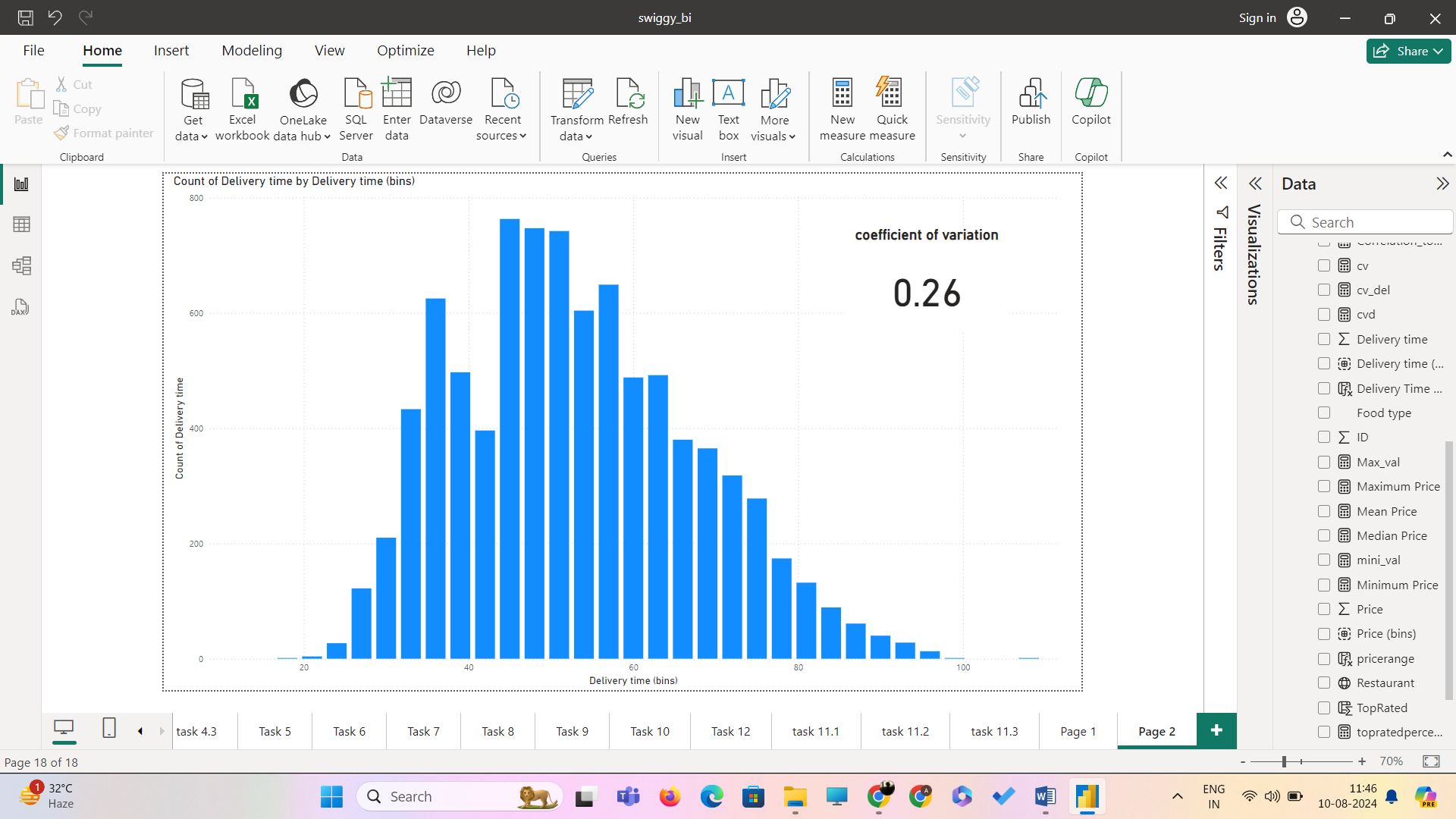
* The mean and median of price suggest the distribution is positively skewed and the clustered column chart named count of price (f) by price(bins) shows the distribution of price and it is skewed right.
* The coefficient of variance (cv)of 0.66 even though less than 1, indicates that the standard deviation is 66% of the mean. This suggests a moderate to high level of relative variability in the dataset. The data points are highly dispersed around the mean and there are outliers present. Further , the spread is strengthened by the standard deviation even though less than mean but it is relatively large.
* In this data set, the standard deviation of 230.93 and central tendency (mean) of 348.44 which explains that most of the data points fall under 100 to 600 and shown in the clustered column chart named count of price (f) by price(bins) which is mean ±1SD.
* The price range of 201-300 occurs the most in the data set.

**Task 8: Delivery Time Analysis**

**Objective**: Analyse the average delivery time of restaurants.



Note: A new measure named delivery time range created

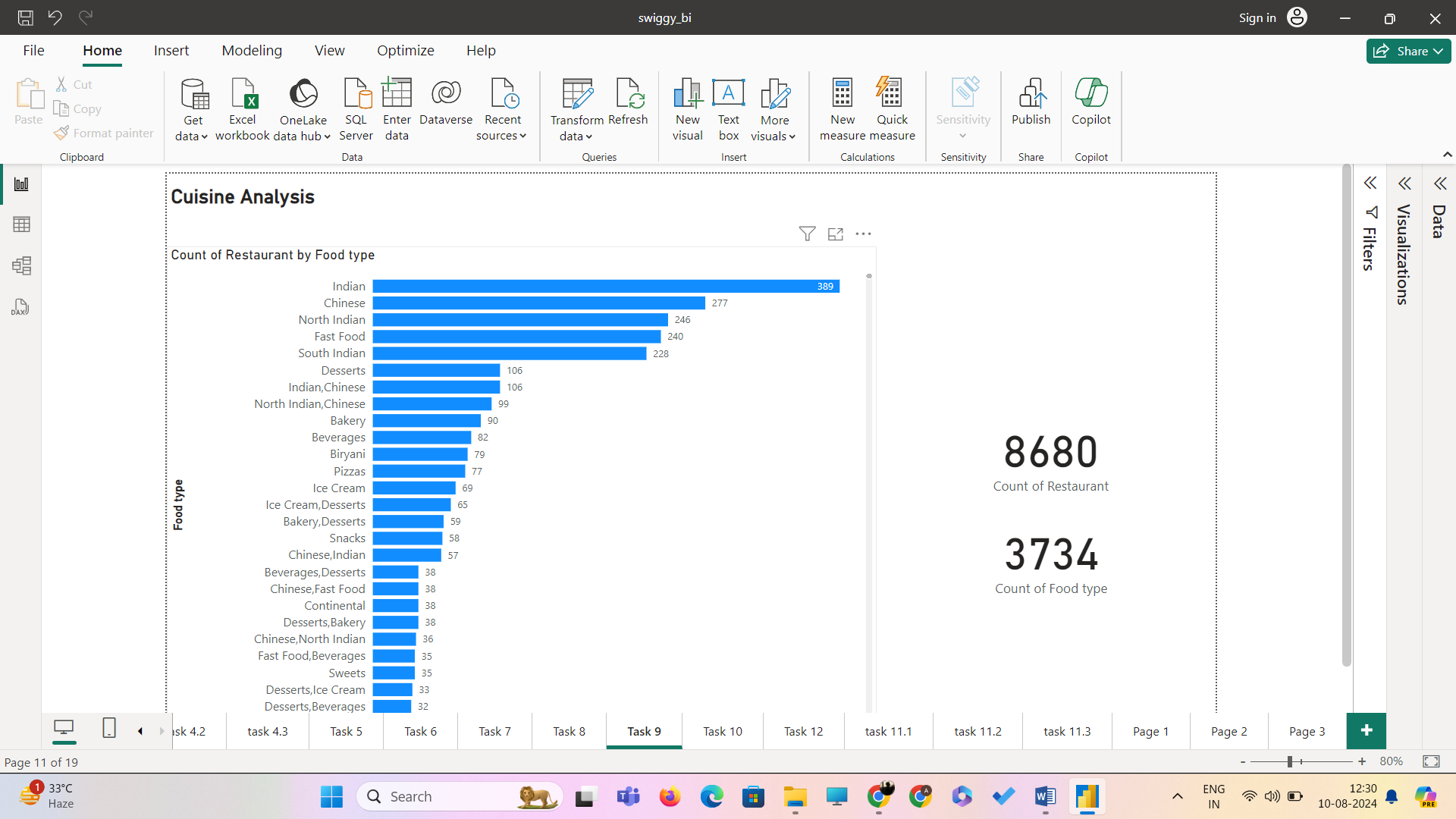


**Analysis and Interpretation:**

* The mean and median of delivery time is almost equal and the difference is only 0.97 but the distribution skewed is slightly right(positive skew). From the cluster column table named count of delivery time by delivery time(bins) shows that the distribution is almost bell-shaped but not a normal distribution.
* The coefficient of variation of delivery time is 0.26 which is 26 % of standard variation which suggests a moderate level of relative variation in the data set . the data points are slightly dispersed around the mean not excessively
* Most of the data points fall in the range of 40-60 which is (mean±1sd).
* The range 41-60 has the highest frequency.

**Task 9: Cuisine Analysis**

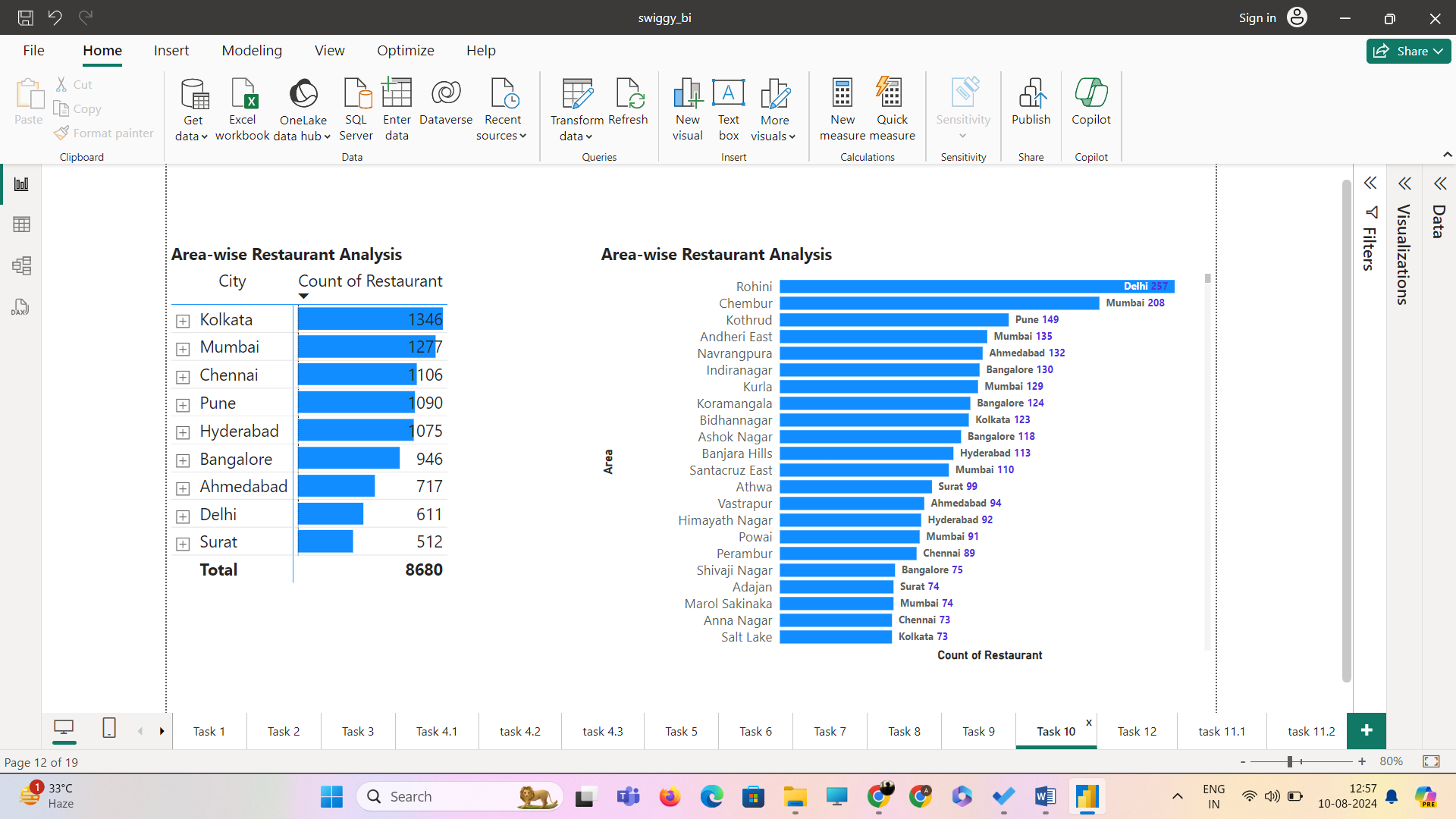
**Objective: Analyse the variety of cuisines offered by restaurants.**



**Interpretation:**

In 8680 Restaurants across India 3734 food types are served and the cuisine served in the highest number of restaurants is Indian.

**Task10: Area-wise Restaurant Analysis**

**Objective**: **Analyse the number of restaurants in each area within the city.**

For area-wise restaurant analysis I have used the matrix table with city as main and area as sub and converted the matrix table into a clustered bar chart. Both table and bar chart can be drilled up and down for future analysis.

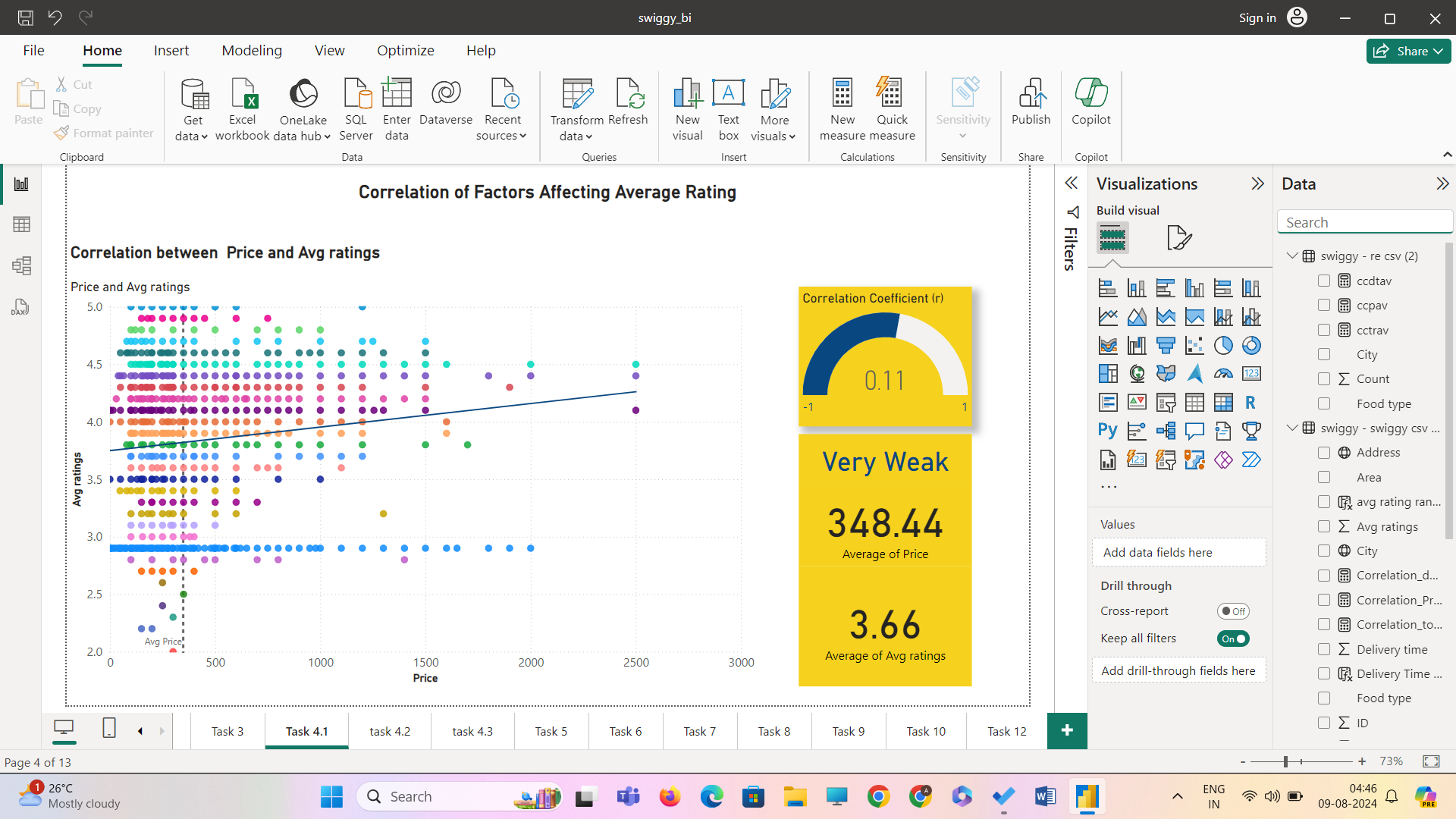
**Interpretation:**

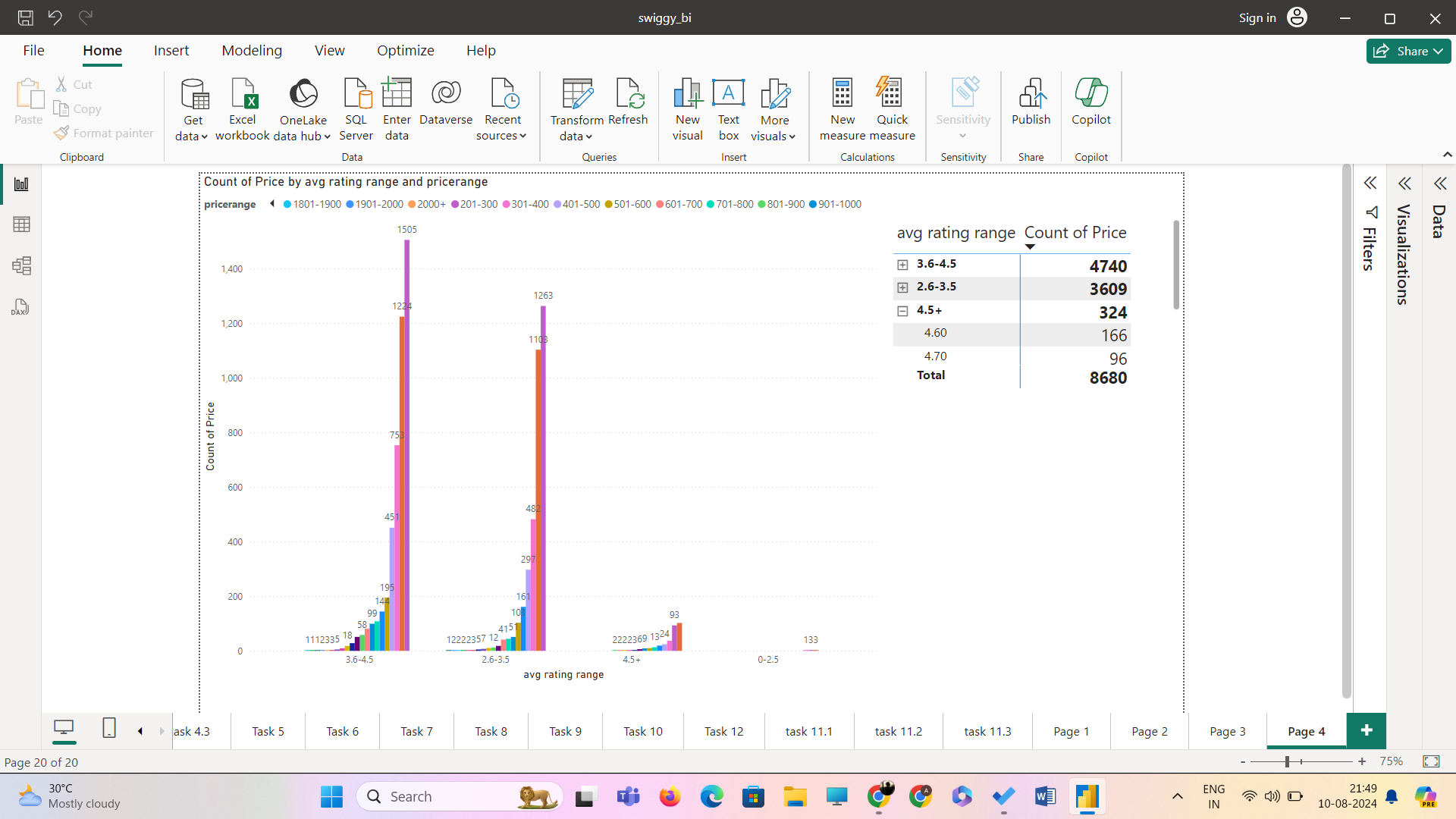
The area with the most restaurants is Rohini.

**Task 11: Correlation Analysis**

**Objective**: **Investigate any correlations between variables such as price, ratings, and delivery time.**

**Correlation between Price and Avg Ratings**





**Interpretation:**

The data points are well dispersed due to right skew of price and left skew of avg ratings and there are outliers in the data set.

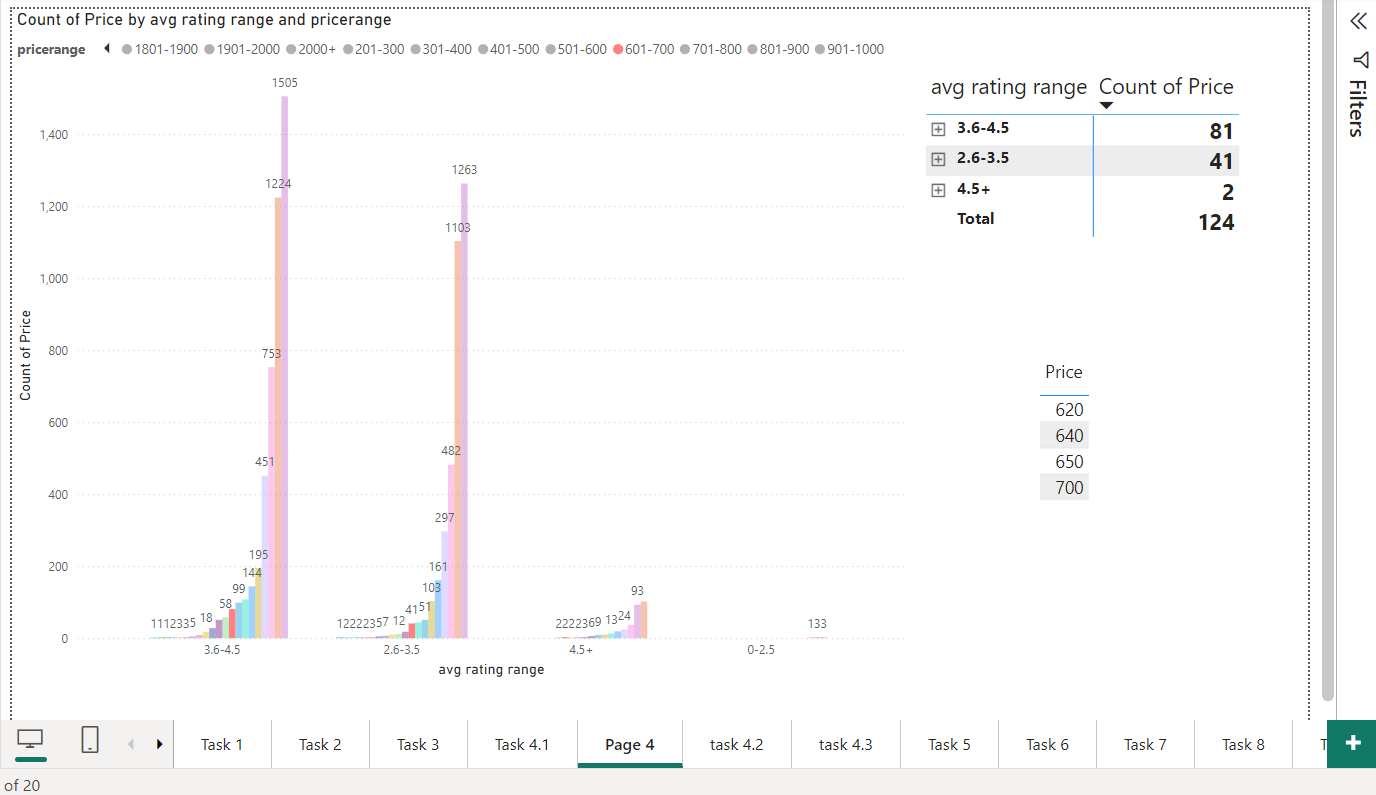
The coefficient correlation( r) = 0.11 explains the relationship between price and avg rating as positive relationship meaning price increases and avg rating also increase but the r value is very small closer zero indicating weak positive correlation making the high priced restaurants may have slightly higher avg rating , but influence of price on ratings is minimal.

The price range of 201-300(below average of price) has the most occurrences in the avg rating range of 2.6-3.5 (below average of avg rating)and 3.6-4.5(above average of avg rating). And the price range of 101-200 has the most occurrence in the avg rating range of 4.5+(above average of average rating) .

The price range of 301-400 where average price falls has third most occurrences in 2.6-3.5 , 3.6-4.5 and 4.5+.

The avg rating range of 0-2.5 has 7 price points which are below average price.

Let take a case of price range 601-700 which above average price



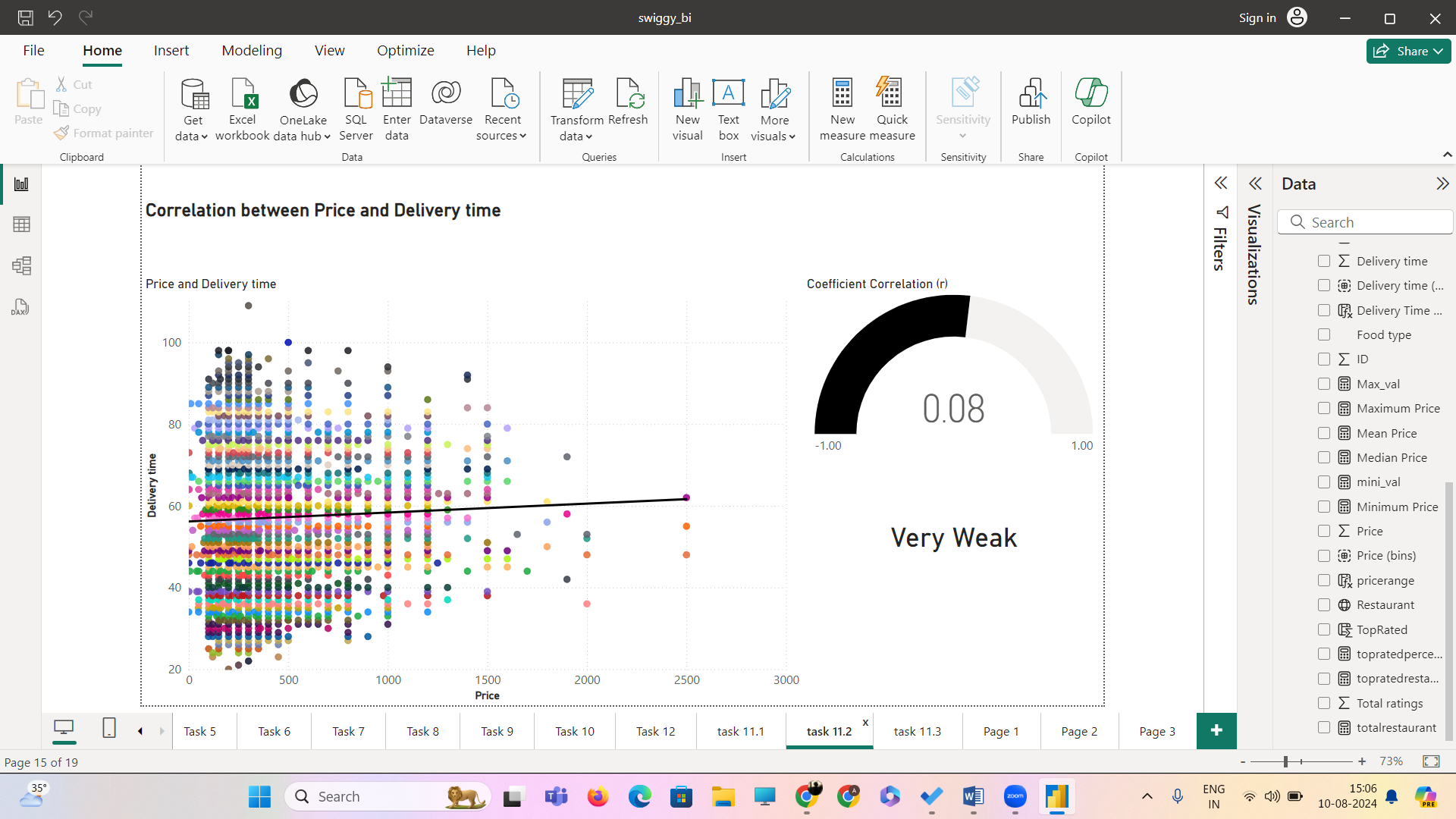
3.6-4.5 avg rating range is where the average of avg ratings falls in that range there is 81 occurrences of price points and in 2.6-4.5 avg rating range 41 occurrences and 4.5+ Avg rating range only 2 occurrences of price points in range of (601-700)

As from the clustered column chart named count of price of avg rating range and price range.

* There are points which are high price with low ratings and low price with high ratings pulling down the positive correlation
* As we are dealing with a strong right skewed distribution(price) 2.6 and left skewed distribution(avg rating)-0.102 as these skews are plotted in a scatter plot and high variance , this suggests non-linear relationship.

The slight positive correlation suggests that there may be a customer perception regarding their preference for high priced restaurants , assuming high price relating to high quality. This also indicates that price is not the only variable affecting avg ratings but there are other factors .

**Correlation between Price and Delivery time**

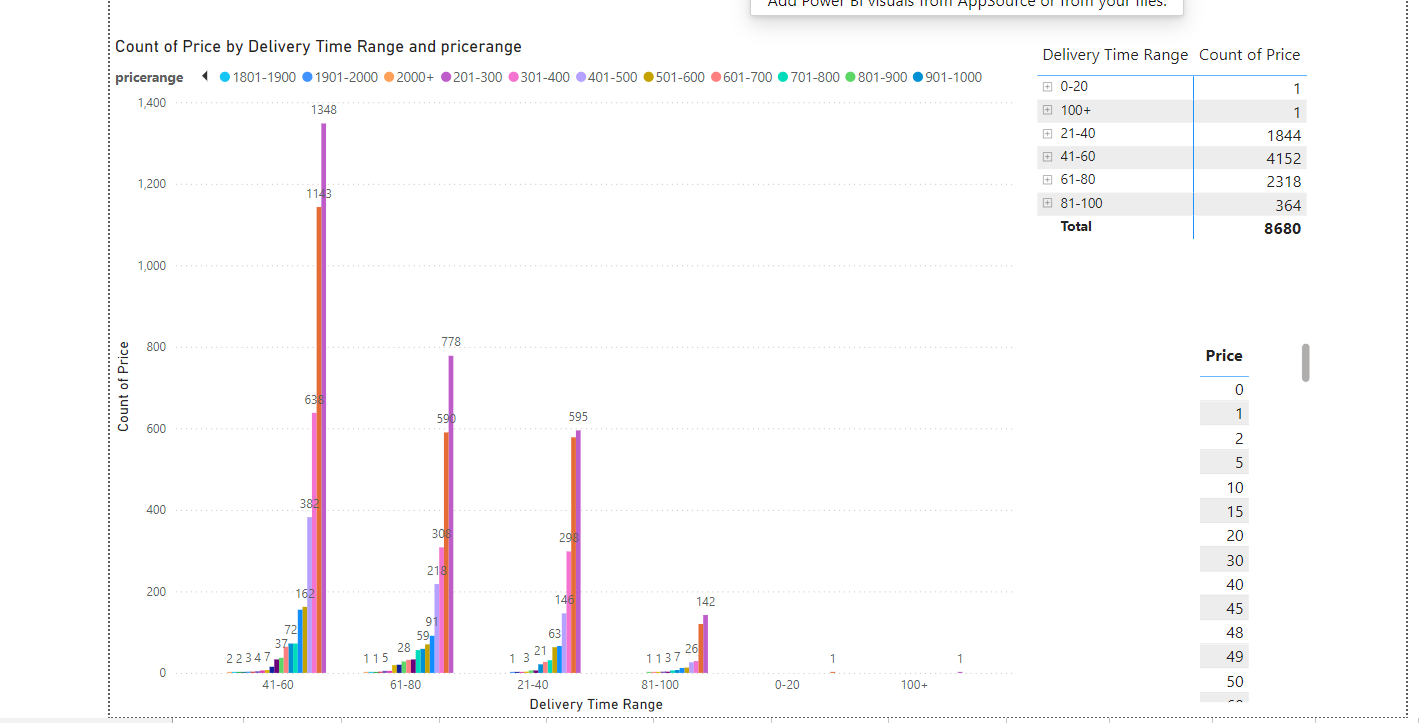
****

**Interpretation:**

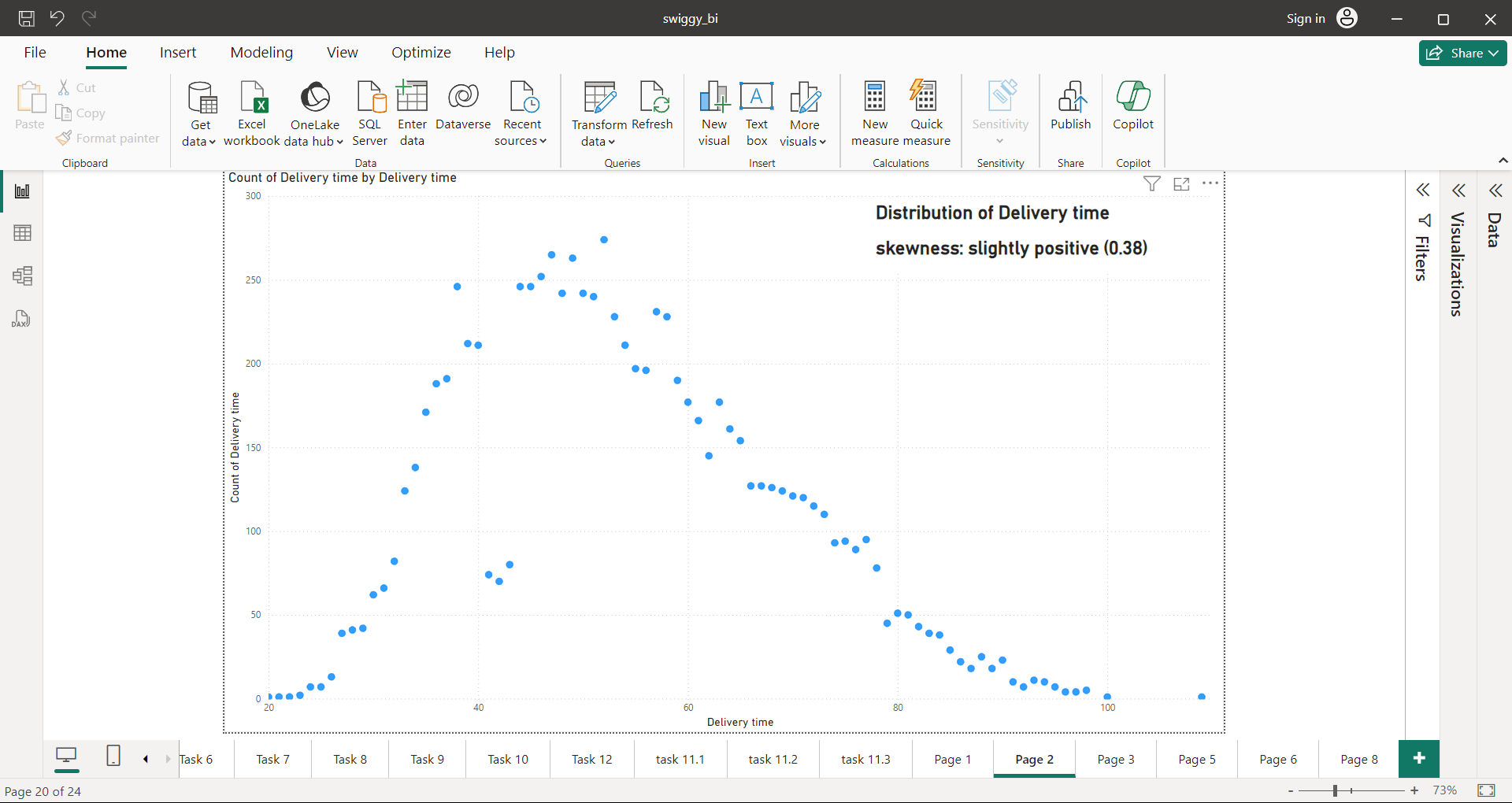
The correlation between price and delivery time is extremely weak , the r value of 0.08 is very close to zero and it can be also taken as negligible and the trend line very slightly faces upwards indicating a very little positive correlation.

The price range of (201-300) has the highest occurrence in in delivery time range of (41-60),(61-80),(21-40)and (81-100)

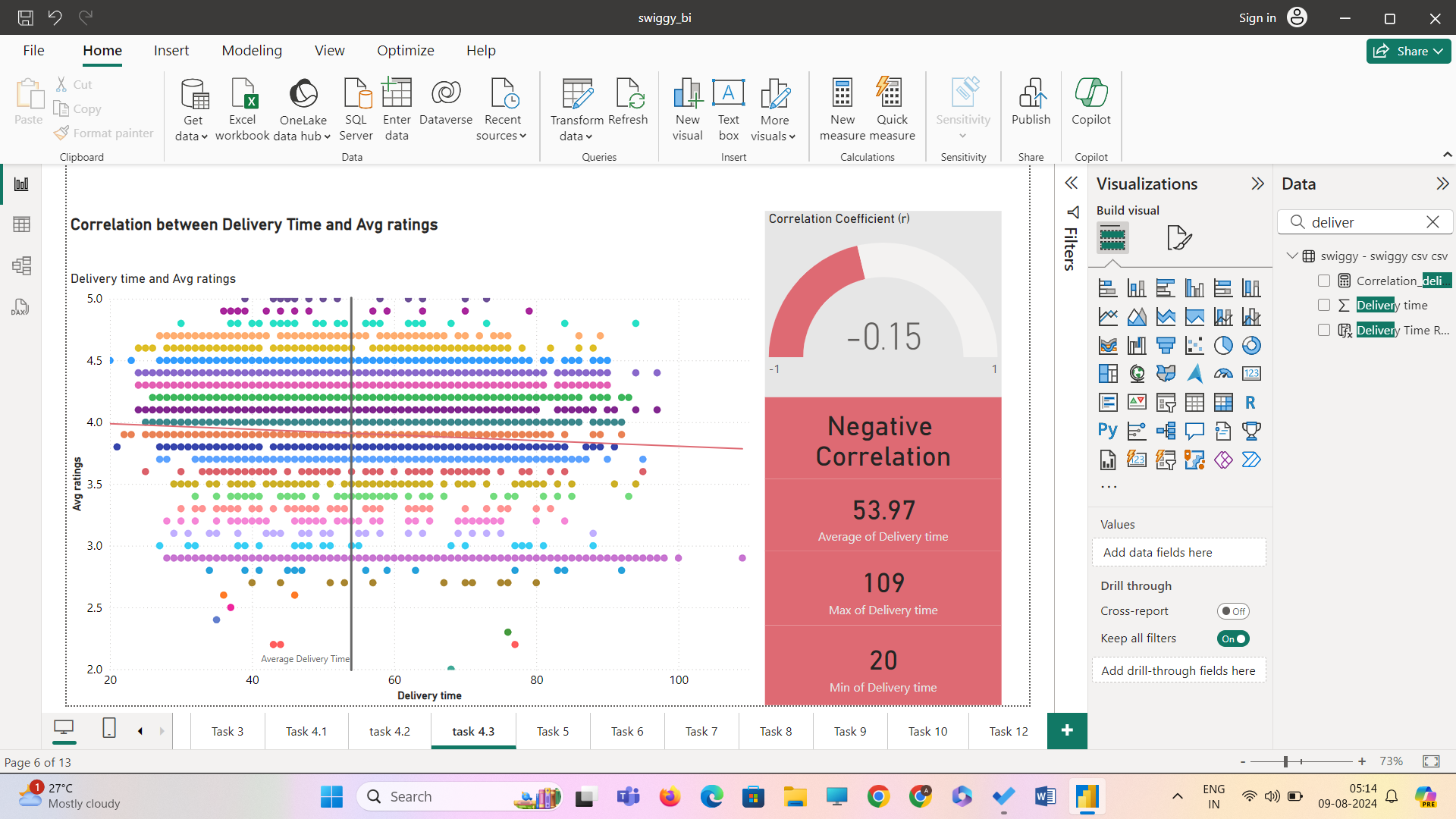
The average delivery time is 53.97 which falls under the range of 41-60, the range 41-60 delivery time condensed data points.



We are dealing with two positive skewed distributions , where there are extreme outliers and with high variance and there is a weak relationship between these as the delivery time and price. As the delivery time increases the price will also increase due to increased distance and high traffic time and vice-versa. There are points where delivery time is high but price is low .



**Correlation between Delivery time and avg Ratings**



**Interpretation:**

The r value is -0.15 which explains that the correlation between delivery time and avg ratings is a very weak negative correlation due to closeness to zero and not dominant.

The dispersion of data points are so high and there are more outliers affecting the trend line and making it weak.

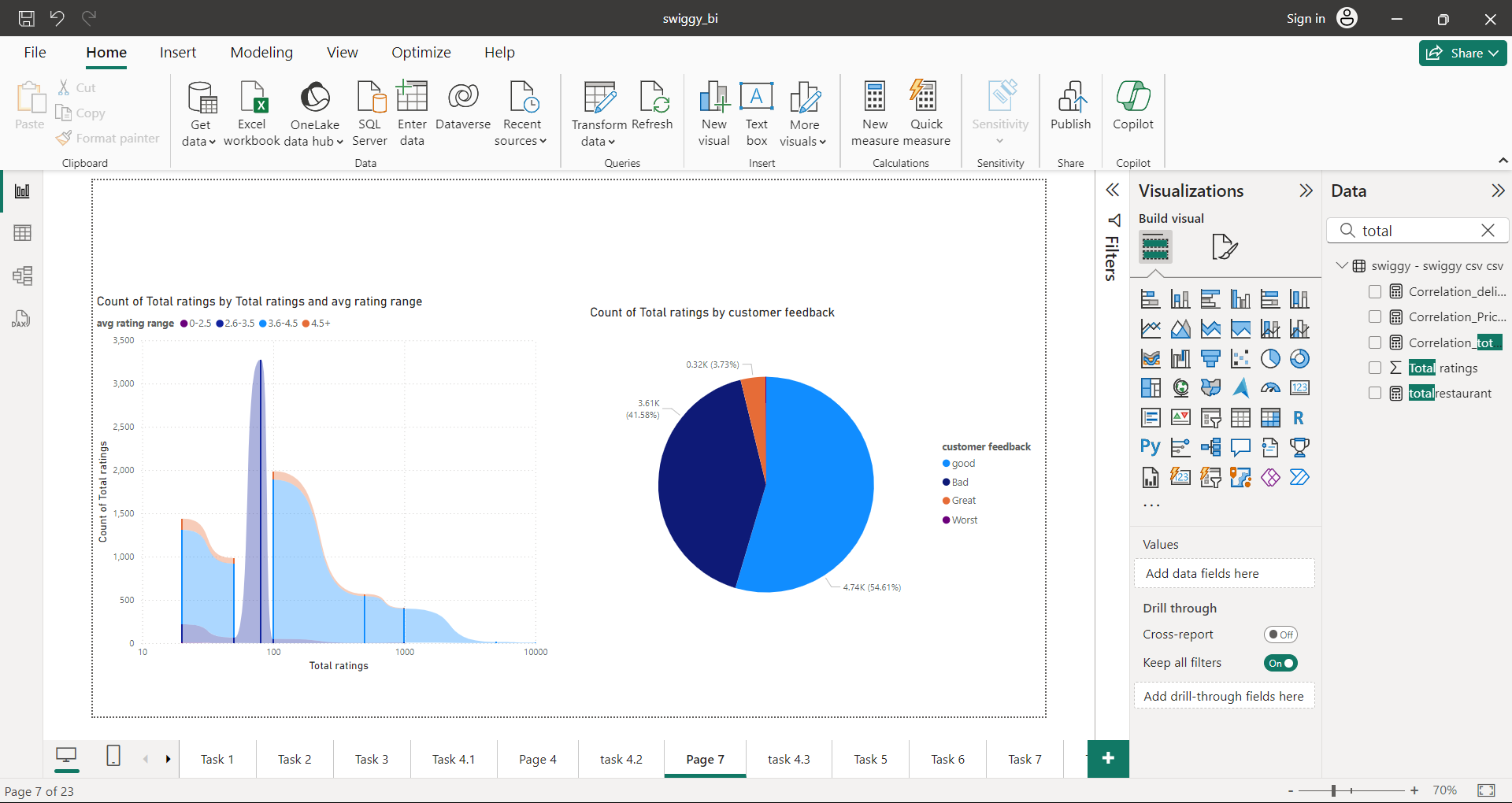
The relationship between these two is a non linear relationship.

The slightly negative correlation may suggest that customers ratings get low as the delivery time increases and vice-versa, but the delivery time as a factor influences avg rating in a very small way.there many other factors such as packaging, value, food taste, estimated time etc also affect the avg ratings of a restaurant.

**The delivery time influences avg ratings than price .**

**Task 12: Customer Feedback Analysis**

**Objective: Analyse customer feedback based on ratings and total ratings.**

****

From the previous task , correlation between total ratings and avg ratings is very weak and the relationship between them is complex.

* 3.73% of the feedback from customers came as great
* 54.61% of the feedback from customer came as good
* 41.58% of the feedback from came as bad
* feedback came as bad from 170 customers.

**Task 14: Business Recommendations**

**Objective**: Provide actionable business recommendations based on the analysis.

* From the analysis , the factors such as price,delivery time, total rating are not only the influencers of avg ratings. Other factors such as quantity and quality of product(food),packaging,perceived value of the service and product, distance , locality and high traffic time (lunch hours) also influence the average ratings.
* In price analysis, it is clear that the points are spread out a lot and have outliers. The reason for these outliers may be influenced by handling fee in high traffic hours and delivery time.as we know that delivery time mostly depend on distance between restaurant and customer location and it increases the price of the order, these are the points where even though delivery time and price is high it has average to above average ratings.
* The problems are only with points where price is high and ratings are bad and worst, the reason behind this may be delivery time exceeds estimated time , packaging issues and quantity and quality of food ,taste.this not only affects restaurants it also affects the image and business of swiggy. It is important to work with feedback and motivate the delivery partners to deliver near estimated time and update the ratings of the restaurant in the application which inturn motivates the business owner to sell good food .
* Swiggy should partner with more restaurants in the cities which might decrease the delivery time giving more options for customers in their locality and keep the handling fee reasonable which may decrease the price.
* spread in price, delivery and avg rating and total rating can also be result by the bullwhip effect caused unregulated discounts,sudden order spike,unavailability of delivery partners and misforcasting. It can be combated by acquiring more swiggy partners and by proper motivation and training can be helpful during sudden order spikes and during high traffic hours. It is important to forecast future order spikes and communicate with restaurants and delivery partners. Providing regulated discounts and communicating with partners and customers about the discount offer details.
* Pushing feedback forms to customers and making it habitual is very important and helps swiggy to be competitive.