EXPLORATORY DATA ANALYIS

SWIGGY

Sales Trends

Home About Contact

INTRODUCTION

OVERVIEW

This project explores the Swiggy dataset to uncover patterns in Delivery Time, Rating, and Sales, leveraging data visualization and statistical analysis

Home About Contact

OBJECTIVES OF THE PROJECT

- Understand the structure and summary of the dataset
- Explore relationships between multiple variables (e.g., Price, Ratings, Delivery Time, and Food Type).
- Compare average ratings or price across cities or areas.
- Summarize key insights like:

1. Which city/area has the best-rated restaurants?
2. Average delivery time by city.
3. Common cuisine types and their ratings.

Home About Contact

THE DATA ANALYSIS PROCESS

Data Collection

Utilized Swiggy dataset containing delivery time, sales, Prices, Area and Rating statistics.

Cleaning

Removed missing values and created new metrics

Data Analysis

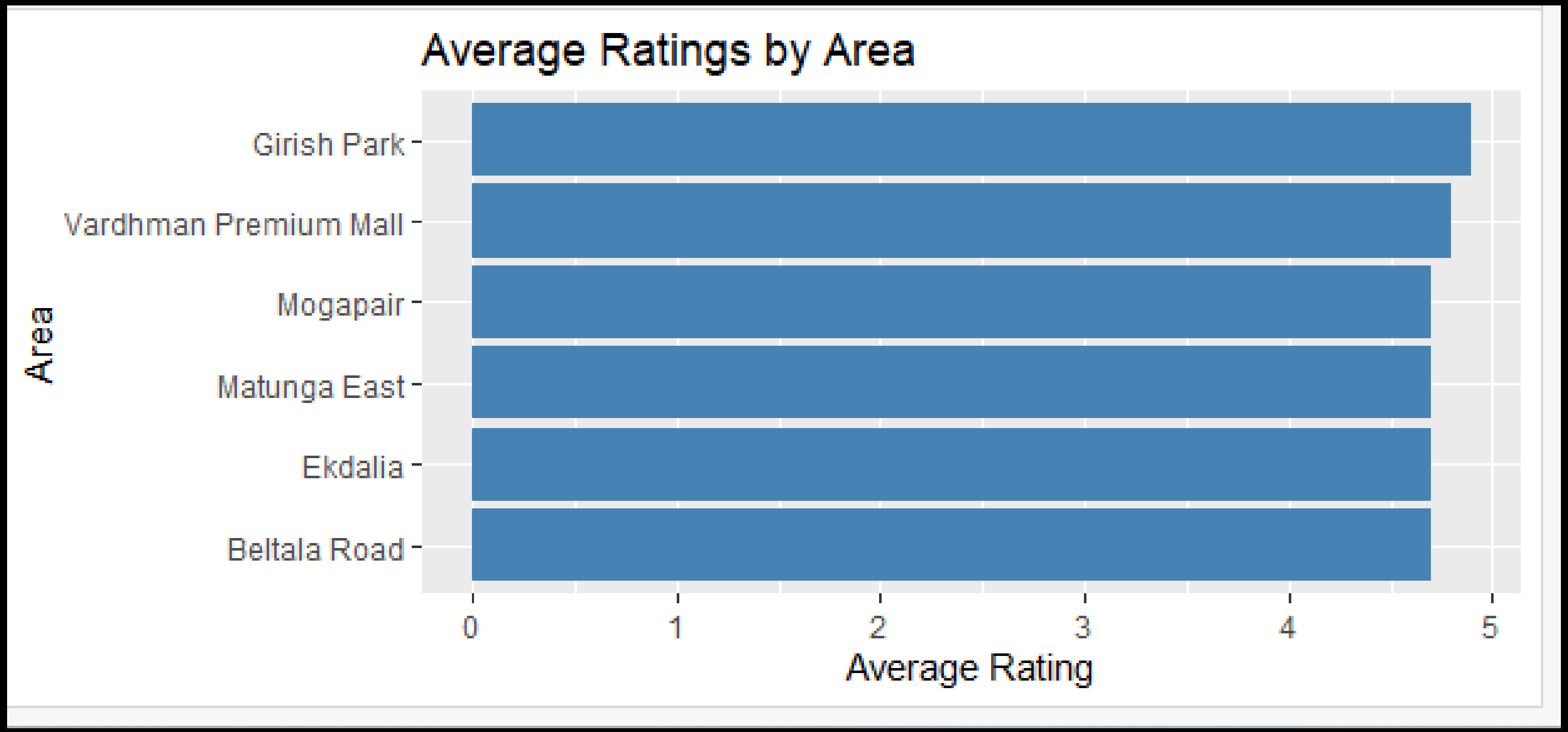
Descriptive Analysis: Summary statistics for Sales, Rating, and Delivery patterns. Visualizations: Bar charts, pie charts, density plots, and box plots for insights. Correlation Analysis: Examined relationships between Area , Rating and Restaurants count

DATA AND VARIABLE

The dataset contains 8,680 rows and 10 columns.

- ID: A unique identifier for each restaurant.
- Area: The locality where the restaurant is situated.
- City: The city in which the restaurant operates.
- Restaurant: Name of the restaurant.
- Price: Average price of a meal or item.
- Average_ratings: The average rating of the restaurant.
- Number_of_Reviews: The number of reviews received by the restaurant.
- Food_Type: The types of cuisines offered by the restaurant.
- Address: Address details of the restaurant.
- Delivery.Time: Estimated delivery time in minutes.

ANALYSIS-1




```
area_avg_ratings <- swiggy_data %>%
  group_by(Area) %>%
  summarize(Average_Rating = mean(Average_ratings)) %>%
  arrange(desc(Average_Rating))
head(area_avg_ratings)

#Viewing the result
View(area_avg_ratings)

ggplot(head(area_avg_ratings), aes(x = reorder(Area, Average_Rating), y =
Average_Rating)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(title = "Average Ratings by Area", x = "Area", y = "Average Rating")
```

ANALYSIS-2

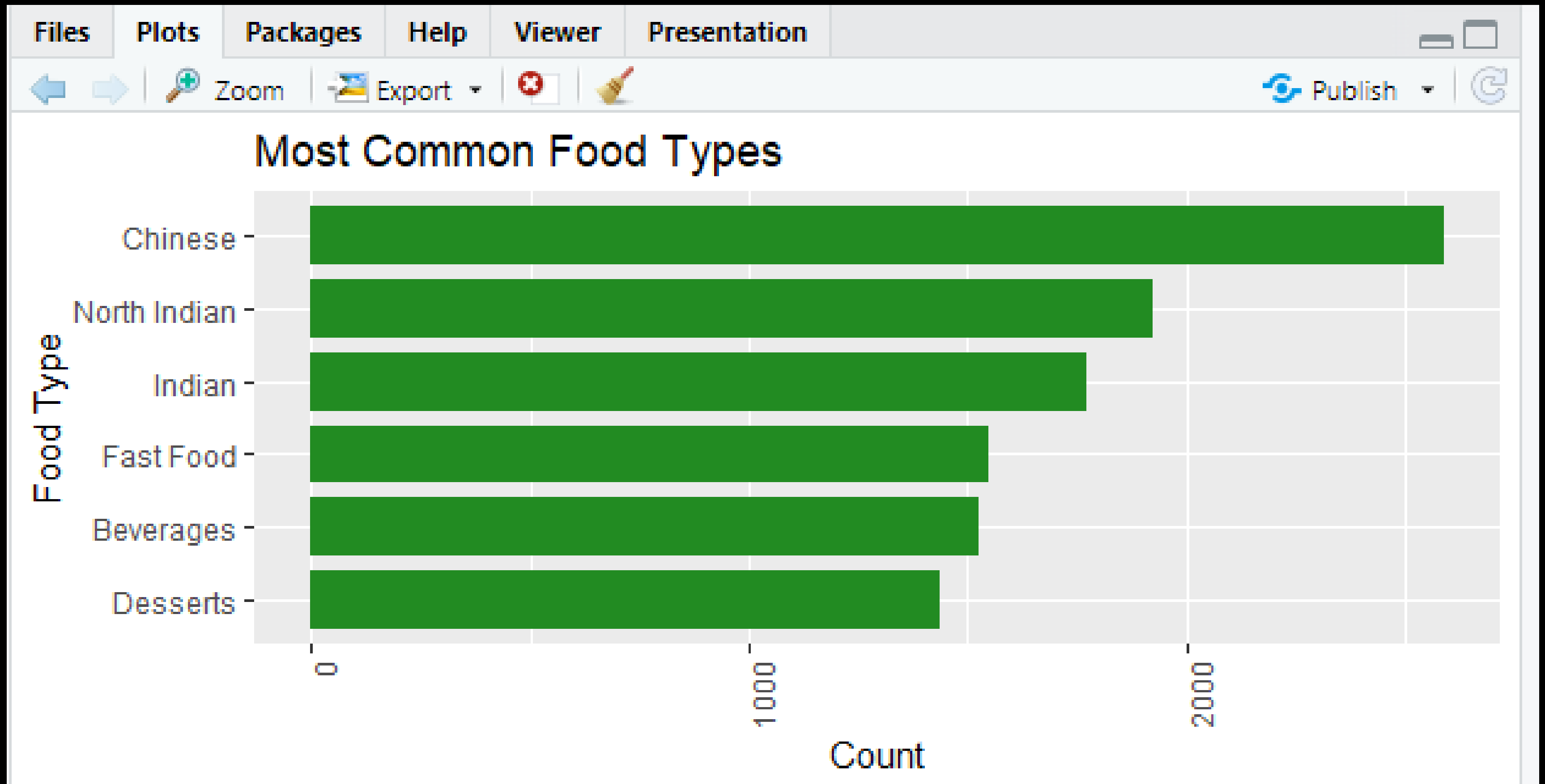
2. What is the distribution of delivery times?

```
ggplot(swiggy_data, aes(x = Delivery.Time)) +
  geom_histogram(binwidth = 5, fill = "red", color = "black") +
  labs(title = "Distribution of Delivery Times", x = "Delivery Time (minutes)", y = "Count")
```

The distribution of delivery times appears to be right-skewed (or positively skewed), with a long tail towards the right. This suggests that most deliveries are completed within a shorter time frame, while a smaller number of deliveries take significantly longer.

ANALYSIS-3

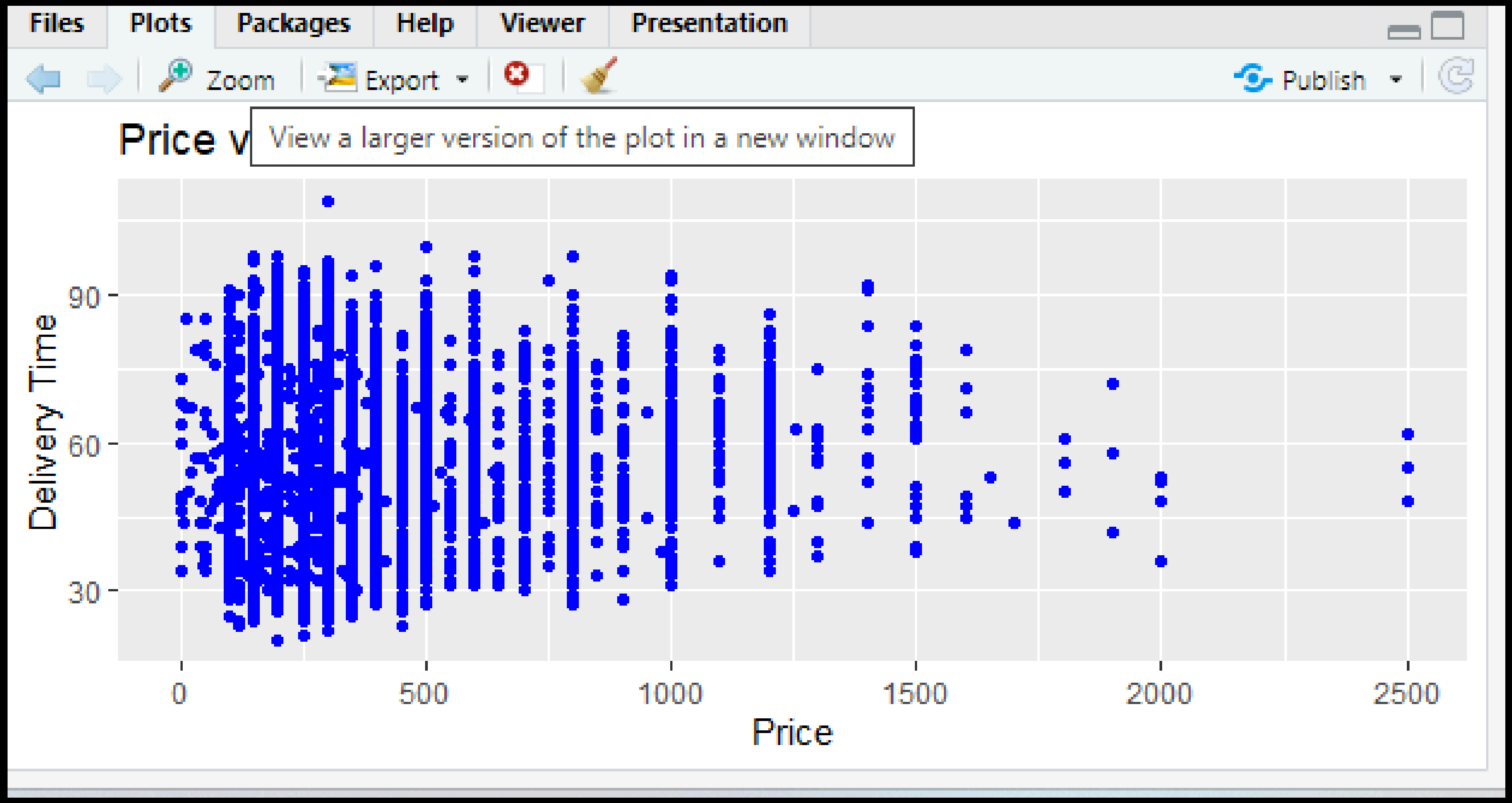
3. Which food types are most commonly offered?




```
food_types <- swiggy_data %>%
  separate_rows(Food_Type, sep = ",") %>%
  count(Food_Type, sort = TRUE)

View(head(food_types))

ggplot(head(food_types), aes(x = reorder(Food_Type, n), y = n)) +
  geom_bar(stat = "identity", fill = "forestgreen", width = 0.8) +
  labs(title = "Most Common Food Types", x = "Food Type", y = "Count") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  coord_flip()
```

ANALYSIS-4

4. What is the relationship between price and delivery time?




```
ggplot(swiggy_data, aes(x = Price, y =
Delivery.Time)) +
  geom_point(color = "blue") +
  labs(title = "Price vs Delivery Time", x = "Price", y =
"Delivery Time")
```

There is no clear linear relationship between Price and Delivery Time. The points are scattered without a distinct trend or pattern.

There seems to be a slight concentration

ANALYSIS-5

5. Which restaurants have the most reviews?

```
restaurant_reviews <- swiggy_data %>%
  arrange(desc(`Number_of_ Reviews`))

View(head(restaurant_reviews))
# Top 10 Restaurants by Review
```

head(restaurant_reviews) × | Gradedassignment_Bhavyanshu.R × | head(food_types) × | area_avg_ratings ×

Filter

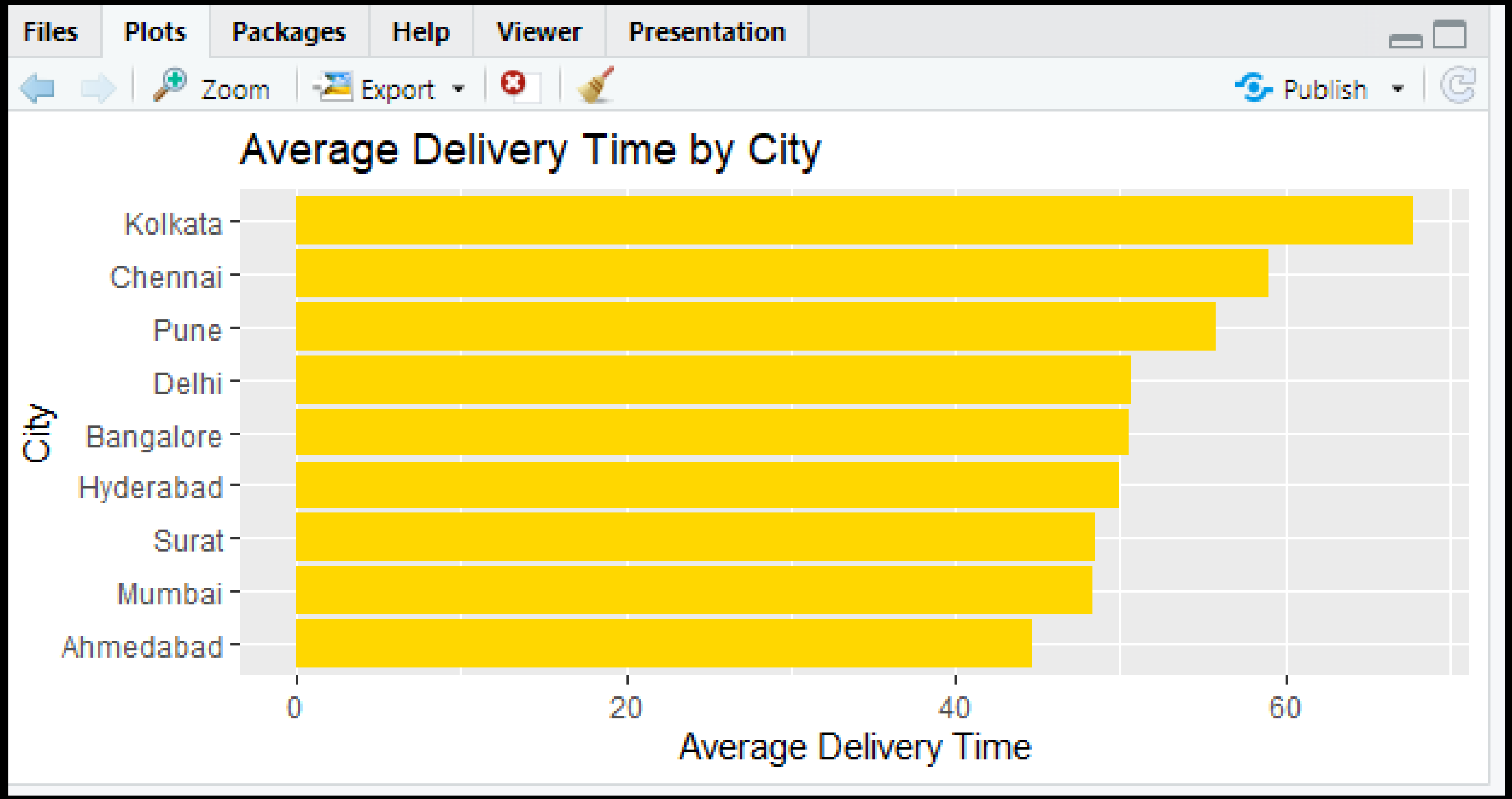
Go back to the previous source location (Ctrl+F9)

					Price	Average_ratings	Number_of_ Reviews	Fo
1	4138	Koti	Hyderabad	Grand Hotel	300	4.1	10000	Bir
2	4320	Narayanguda	Hyderabad	Mehfil	300	3.9	10000	Bir
3	4650	New Nallakunta	Hyderabad	Bawarchi	300	3.9	10000	Bir
4	34634	Toli Chowki	Hyderabad	Shah Ghouse Cafe & Restaurant	350	4.0	10000	Bir
5	131870	Kavadiguda	Hyderabad	4M Biryani House	250	4.1	10000	Bir
6	367769	Chintal	Hyderabad	Lucky Restaurant	300	4.0	10000	Bir

ANALYSIS-6

6. What is the average delivery time by city?

```
city_avg_delivery <- swiggy_data %>%
  group_by(City) %>%
  summarize(Average_Delivery_Time = mean(Delivery.Time)) %>%
  arrange(desc(Average_Delivery_Time))

#viewing the result
View(city_avg_delivery)

ggplot(city_avg_delivery, aes(x = reorder(City, Average_Delivery_Time), y = Average_Delivery_Time)) +
  geom_col(fill = "gold") +
  coord_flip() +
  labs(title = "Average Delivery Time by City", x = "City", y = "Average Delivery Time")
```

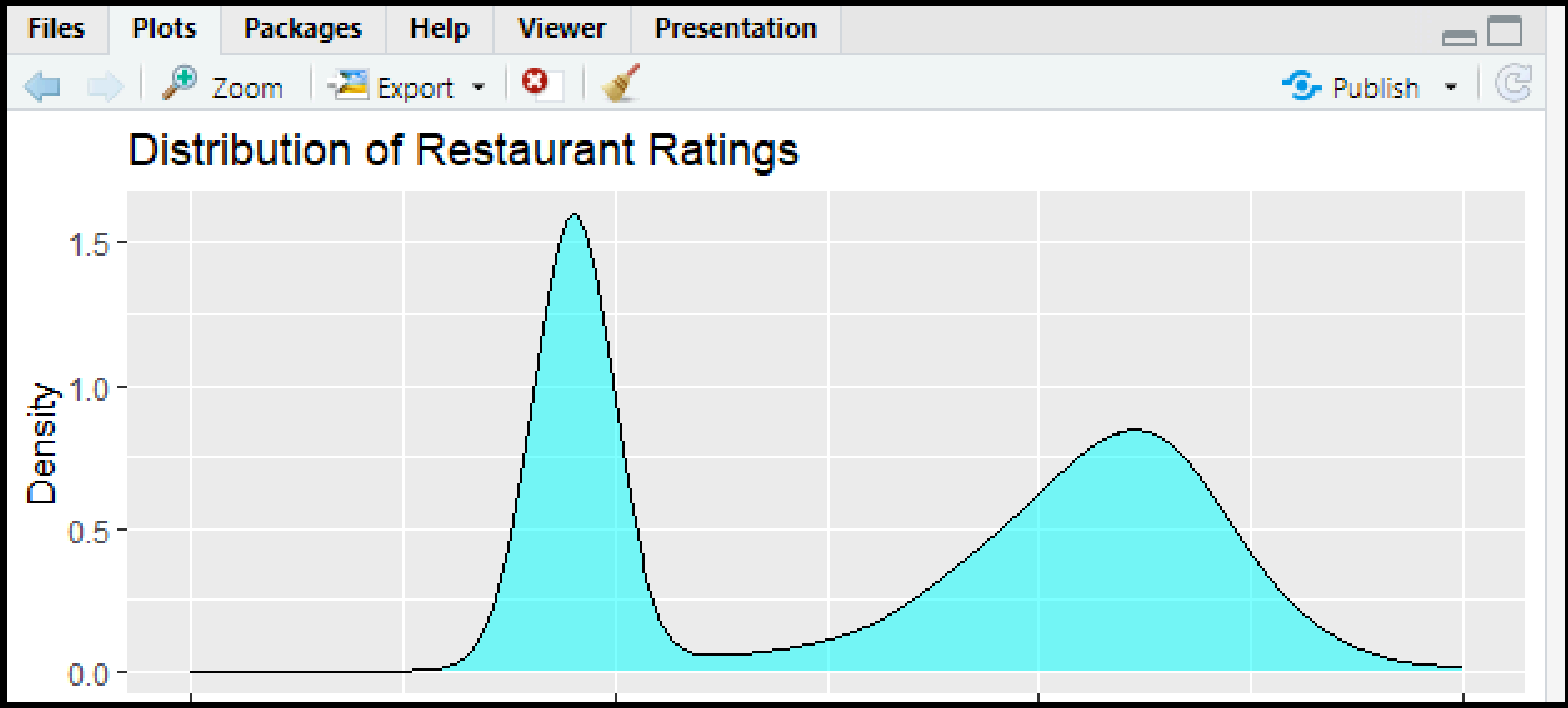


Key Observations: Kolkata appears to have the highest average delivery time, while Ahmedabad has the lowest. There's a noticeable variation in average delivery times across the cities.

ANALYSIS-7

7. What is the distribution of restaurant ratings?

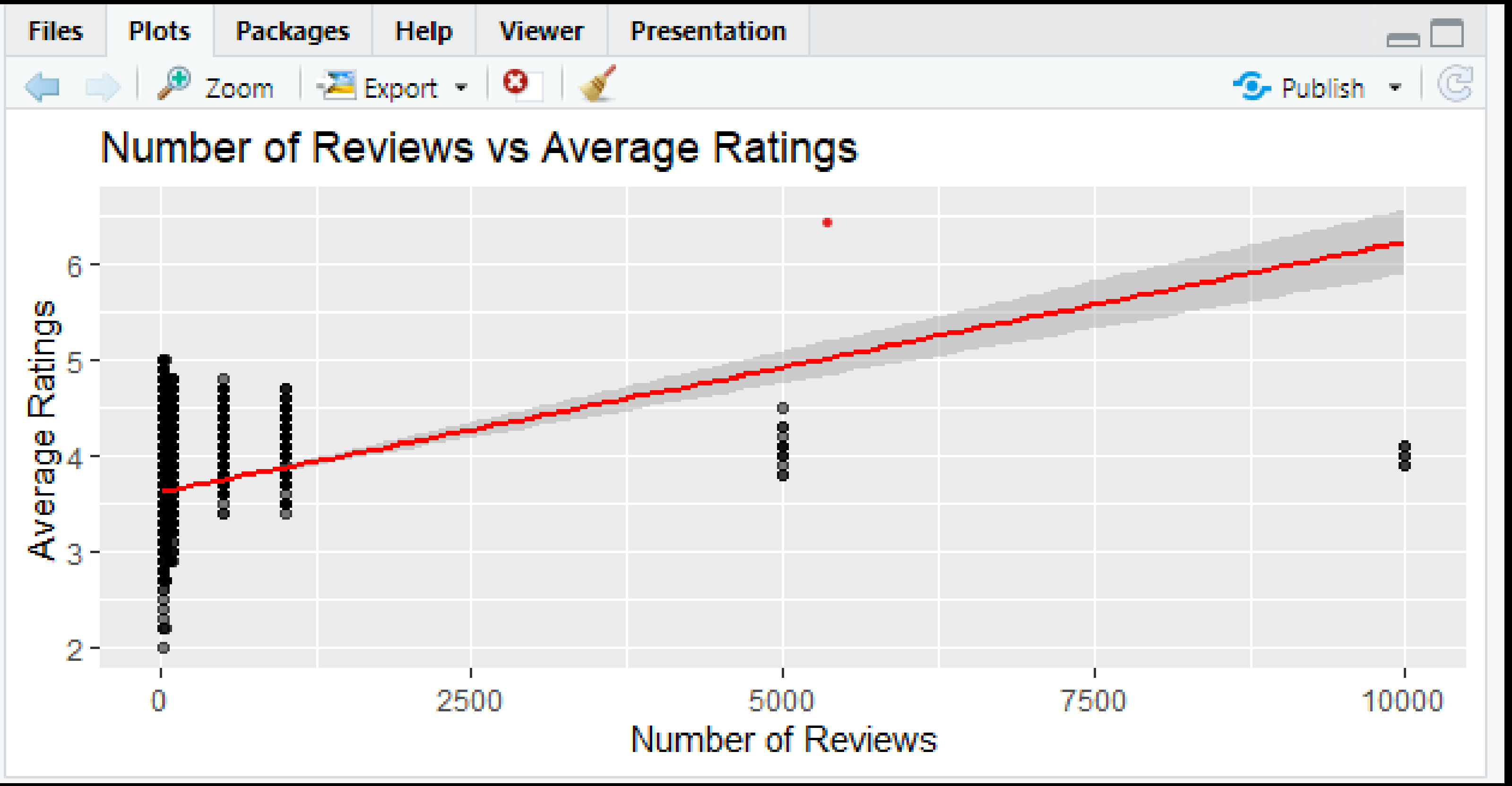
```
ggplot(swiggy_data, aes(x = Average_ratings)) +
  geom_density(fill = "cyan", alpha = 0.5) +
  labs(title = "Distribution of Restaurant Ratings", x =
"Average Ratings", y = "Density")
```



- Key Observations: The density plot shows a bimodal distribution of restaurant ratings. There are two distinct peaks, suggesting that there are two clusters of restaurants with different average ratings.
- Insights: The bimodal distribution indicates that there might be two distinct groups of restaurants in terms of their average ratings. This could be due to factors such as cuisine type, price range, location, or service quality.

ANALYSIS-8

8. How does the number of reviews relate to average ratings?



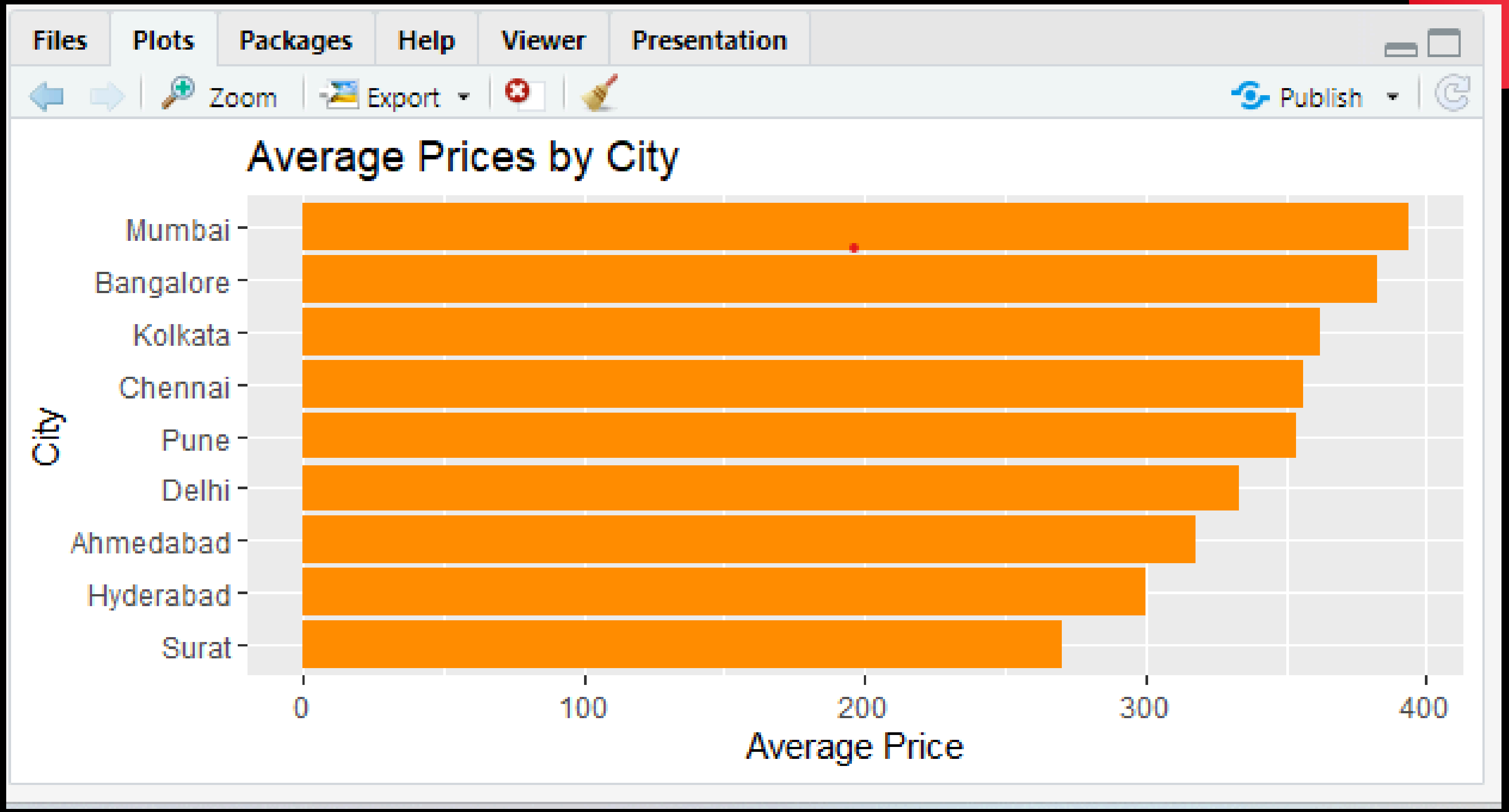

```
ggplot(swiggy_data, aes(x = `Number_of_ Reviews`, y =
Average_ratings)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", color = "red") +
  labs(title = "Number of Reviews vs Average Ratings", x =
"Number of Reviews", y = "Average Ratings")
```

- Key Observations: There appears to be a weak positive correlation between the number of reviews and average ratings. As the number of reviews increases, the average rating tends to increase slightly. However, there's considerable variability in ratings even for restaurants with a high number of reviews.
- Insights: The plot suggests that while more reviews might generally lead to higher average ratings, other factors likely play a significant role in determining a restaurant's rating.

ANALYSIS-9

9. Which cities have the highest average prices?

```
city_avg_price <- swiggy_data %>%
  group_by(City) %>%
  summarize(Average_Price = mean(Price)) %>%
  arrange(desc(Average_Price))

#viewing the result
print(city_avg_price)

ggplot(city_avg_price, aes(x = reorder(City, Average_Price),
y = Average_Price)) +
  geom_col(fill = "darkorange") +
  coord_flip() +
  labs(title = "Average Prices by City", x = "City", y = "Average
Price")
```



- Key Observations: Mumbai has the highest average price, while Surat has the lowest. There's a noticeable variation in average prices across the cities.
- Insights: This chart provides a quick visual comparison of price levels across different locations. It could be used to identify cities with higher or lower average prices, which might be relevant for various purposes such as market research, business planning, or consumer decision-making.

ANALYSIS-10

10. What are the most expensive restaurants?

Gradedassignment_Bhavyanshu.R | expensive_restaurants | city_avg_delivery | head(food_types)

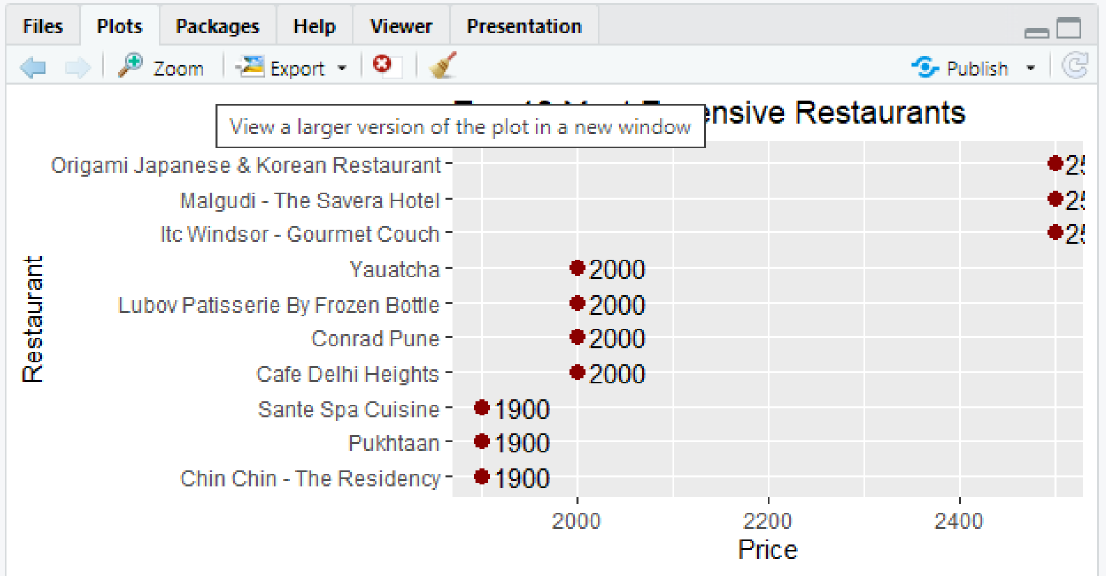
	ID	Area	City	Restaurant	Price	Average_ratings	Number_of_ Reviews
1	294499	Vasanth Nagar	Bangalore	Itc Windsor - Gourmet Couch	2500	4.5	
2	306434	Mylapore	Chennai	Bangalore he Savera Hotel	2500	4.1	
3	394904	Scruz Bandra East	Mumbai	Origami Japanese & Korean Restaurant	2500	4.4	
4	72003	Saki Naka	Mumbai	Cafe Delhi Heights	2000	2.9	
5	266184	Bandra Area	Mumbai	Yauatcha	2000	4.5	1
6	272388	Sangamvadi	Pune	Conrad Pune	2000	4.4	
7	439628	Brigade Road	Bangalore	Lubov Patisserie By Frozen Bottle	2000	2.9	
8	34312	T. Nagar	Chennai	Chin Chin - The Residency	1900	4.3	
9	215427	Kurla	Mumbai	Pukhtaan	1900	2.9	
10	442443	Bandra West	Mumbai	Sante Spa Cuisine	1900	2.9	

Showing 1 to 10 of 10 entries, 10 total columns

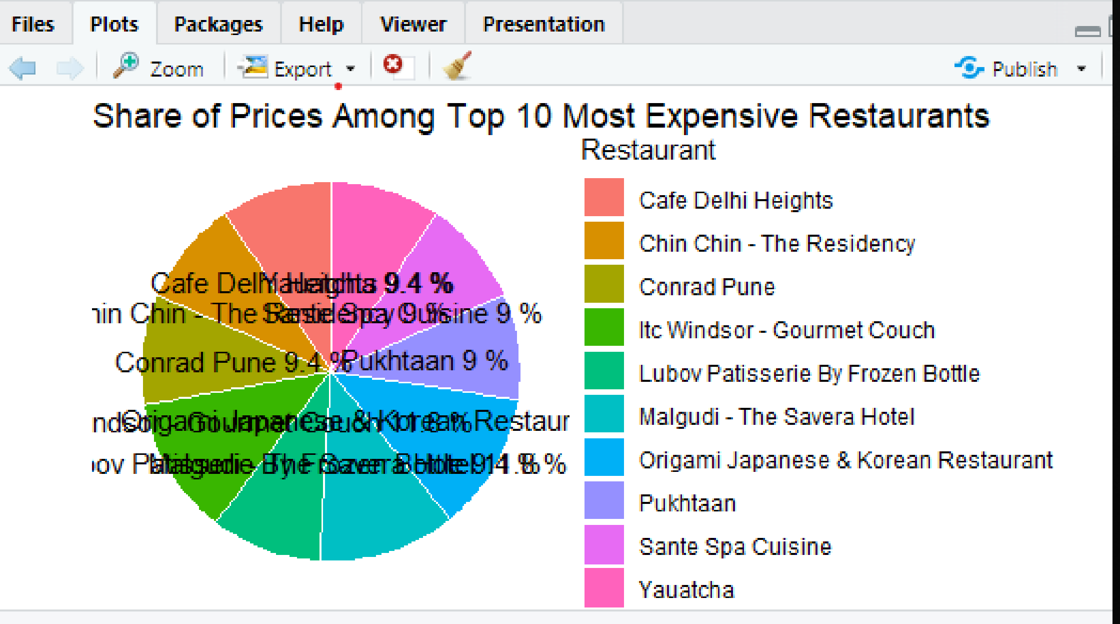
```
expensive_restaurants <-
swiggy_data %>%
  arrange(desc(Price)) %>%
  head(10)

#viewing the result
View(expensive_restaurants)
```

- Key Observations: The shares of prices are relatively evenly distributed among the top 10 restaurants. No single restaurant dominates the price share.
- Insights: This pie chart provides a quick visual comparison of the price distribution among the top 10 most expensive restaurants. It can be used to understand which restaurants have a higher share of the overall price range.

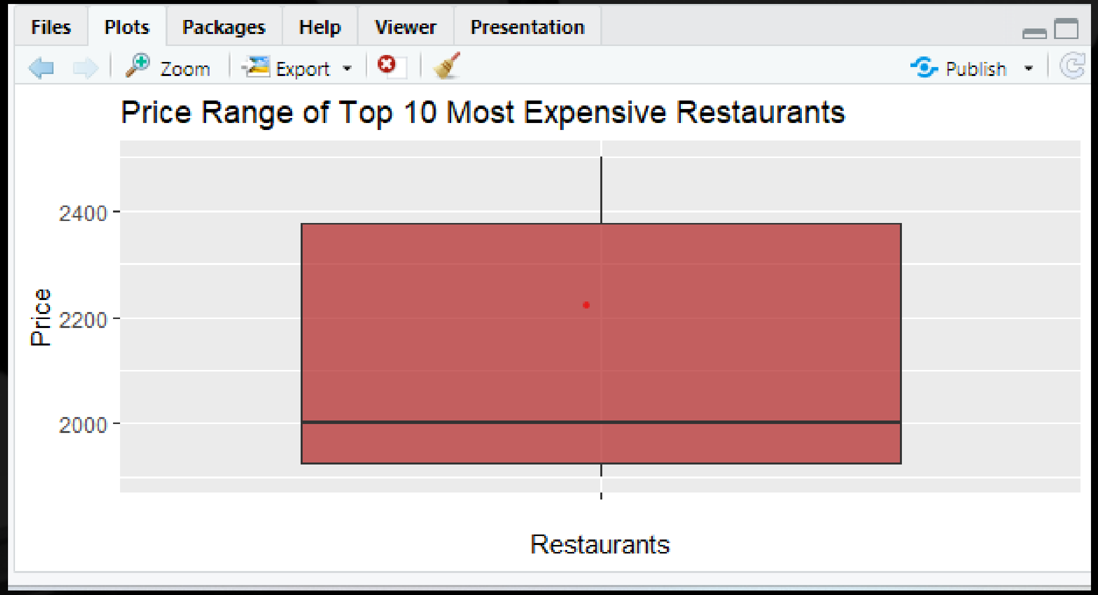
Scatter plot



Pie Chart



Boxplot

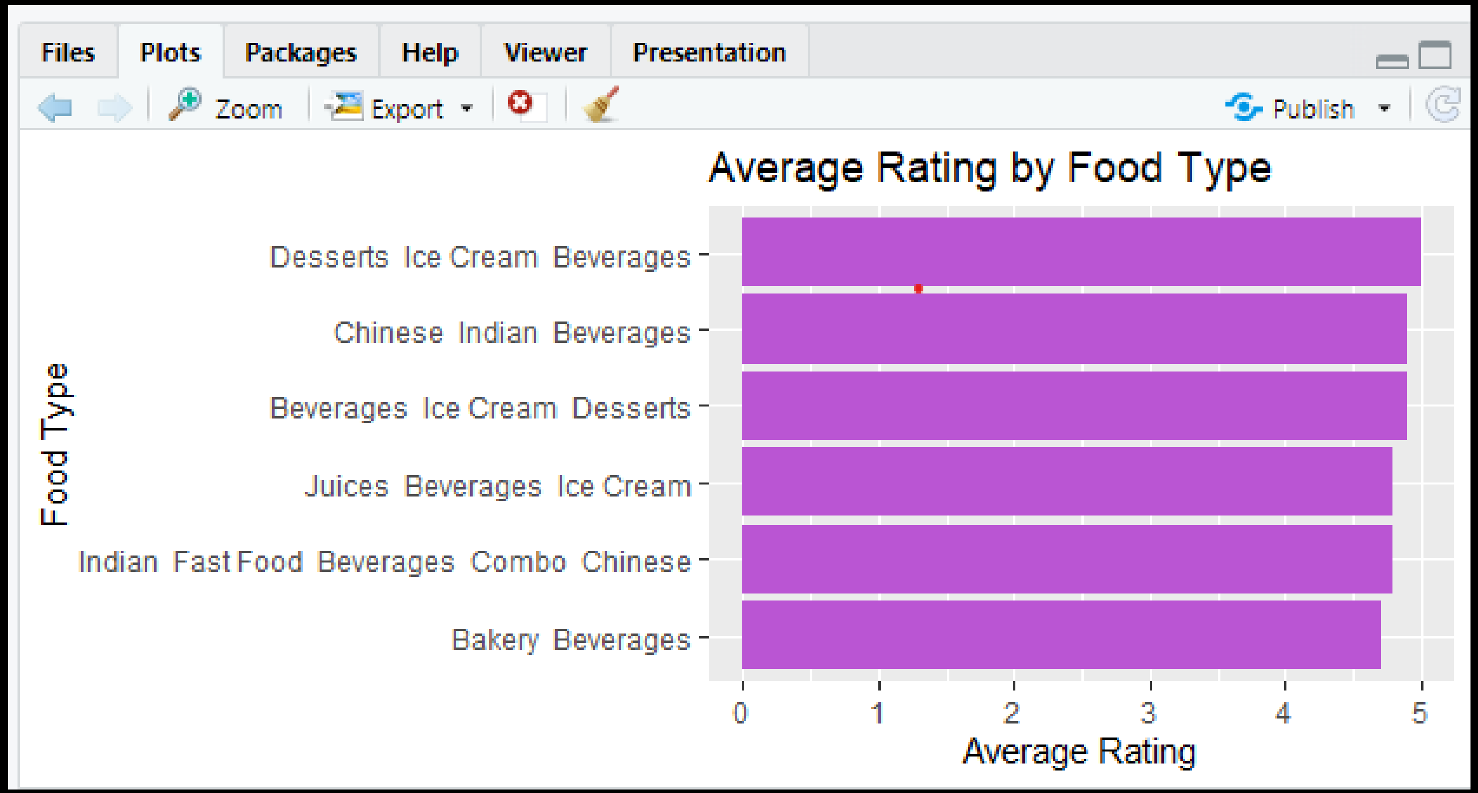


ANALYSIS-11

11. How does average rating vary by food type?

```
food_avg_rating <- swiggy_data %>%
  separate_rows(Food_Type, sep = ",") %>%
  group_by(Food_Type) %>%
  summarize(Average_Rating = mean(Average_ratings)) %>%
  arrange(desc(Average_Rating))
```

```
ggplot(head(food_avg_rating), aes(x = reorder(Food_Type, Average_Rating), y =
Average_Rating)) +
  geom_col(fill = "mediumorchid") +
  coord_flip() +
  labs(title = "Average Rating by Food Type", x = "Food Type", y = "Average Rating")
```



Gradedassignment_Bhavyanshu.R* | food_avg_rating | expensive_restaurants

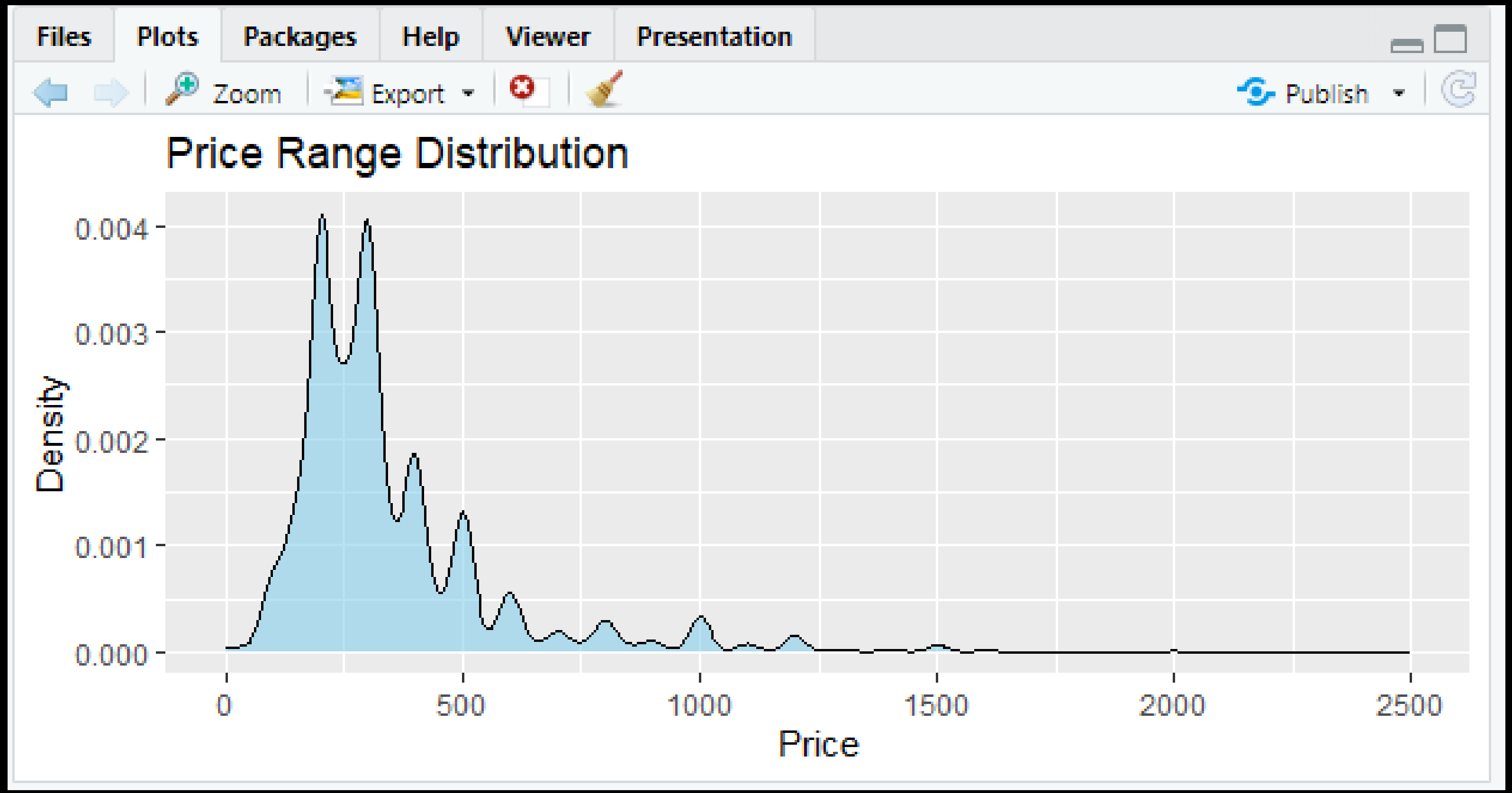
C:/Users/bhavy/OneDrive/Desktop/MABS/GT-BA/R/Gradedassignment_Bhavyanshu/Gradeda

	Food_Type	Average_Rating
1	Desserts Ice Cream Beverages	5.000
2	Beverages Ice Cream Desserts	4.900
3	Chinese Indian Beverages	4.900
4	Indian Fast Food Beverages Combo Chinese	4.800
5	Juices Beverages Ice Cream	4.800
6	Bakery Beverages	4.700
7	Bakery Continental Salads Pastas European Desserts	4.700
8	Cafe Italian Pizzas Fast Food Combo Beverages Desserts	4.700
9	Continental Healthy Food Indian	4.700
10	Ice Cream Cakes Desserts Bakery Sweets Beverages	4.700
11	Sweets Desserts Snacks Home Food Indian	4.700

- Key Observations: The chart shows a variation in average ratings across different food types. "Desserts Ice Cream Beverages" and "Bakery Beverages" appear to have the highest average ratings, while other food types have lower average ratings.
- Insights: This chart provides a quick visual comparison of the average ratings for different food types. It could be used to identify the most popular food categories based on customer ratings, which might be valuable information for menu planning, marketing, or product development.

ANALYSIS-12

12. What is the price range distribution?




```
# Visualization with density plot
ggplot(swiggy_data, aes(x = Price)) +
 geom_density(fill = "skyblue", alpha = 0.6) +
 labs(title = "Price Range Distribution", x = "Price", y = "Density")
```

- Key Observations: The density plot shows a multimodal distribution of prices. There are several distinct peaks, suggesting that there are multiple clusters of prices with different frequencies.
- Insights: The multimodal distribution indicates that there might be different price segments or categories within the data. This could be due to factors such as product type, brand, quality, or target audience.

ANALYSIS-13

13. What is the average delivery time for different food types?

```
food_delivery_time <- swiggy_data %>%
  separate_rows(Food_Type, sep = ",") %>%
  group_by(Food_Type) %>%
  summarize(Average_Delivery_Time =
mean(Delivery.Time)) %>%
  arrange(desc(Average_Delivery_Time))

#viewing the result
View(food_delivery_time)
```

Gradedassignment_Bhavyanshu.R* | food_delivery_time | food_avg_rating | expensive_restaurants

Filter

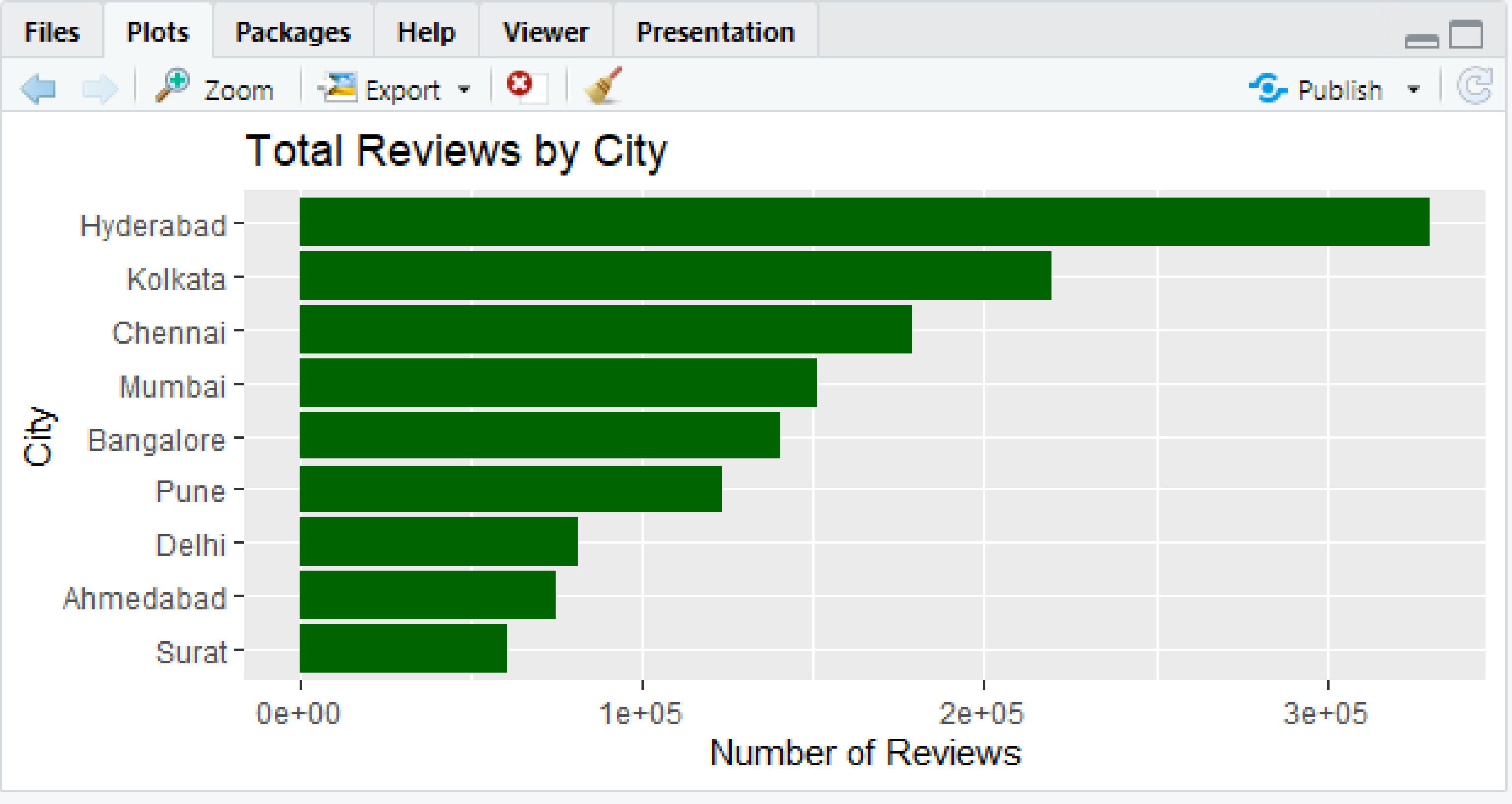
	Food_Type	Average_Delivery_Time
1	Beverages Italian	93.00000
2	Indian Chinese Snacks	88.00000
3	Indian Chinese South Indian Bengali Snacks	88.00000
4	Bengali Chinese	87.00000
5	Desserts Bakery Snacks	87.00000
6	Snacks Beverages	87.00000
7	Indian Continental Beverages Fast Food	86.00000
8	American Fast Food Italian Mexican Beverages	85.00000
9	Bengali Indian Combo	83.00000
10	Healthy Food Juices Beverages Snacks Fast Food	83.00000
11	Indian Fast Food Beverages	83.00000

Showing 1 to 11 of 601 entries, 2 total columns

ANALYSIS-14

14. How does the number of reviews vary by city?

```
city_reviews <- swiggy_data %>%
  group_by(City) %>%
  summarize(Total_Reviews = sum(`Number_of_ Reviews`)) %>%
  arrange(desc(Total_Reviews))

#viewing the result
View(city_reviews)

ggplot(city_reviews, aes(x = reorder(City, Total_Reviews), y =
Total_Reviews)) +
  geom_col(fill = "darkgreen") +
  coord_flip() +
  labs(title = "Total Reviews by City", x = "City", y = "Number of
Reviews")
```

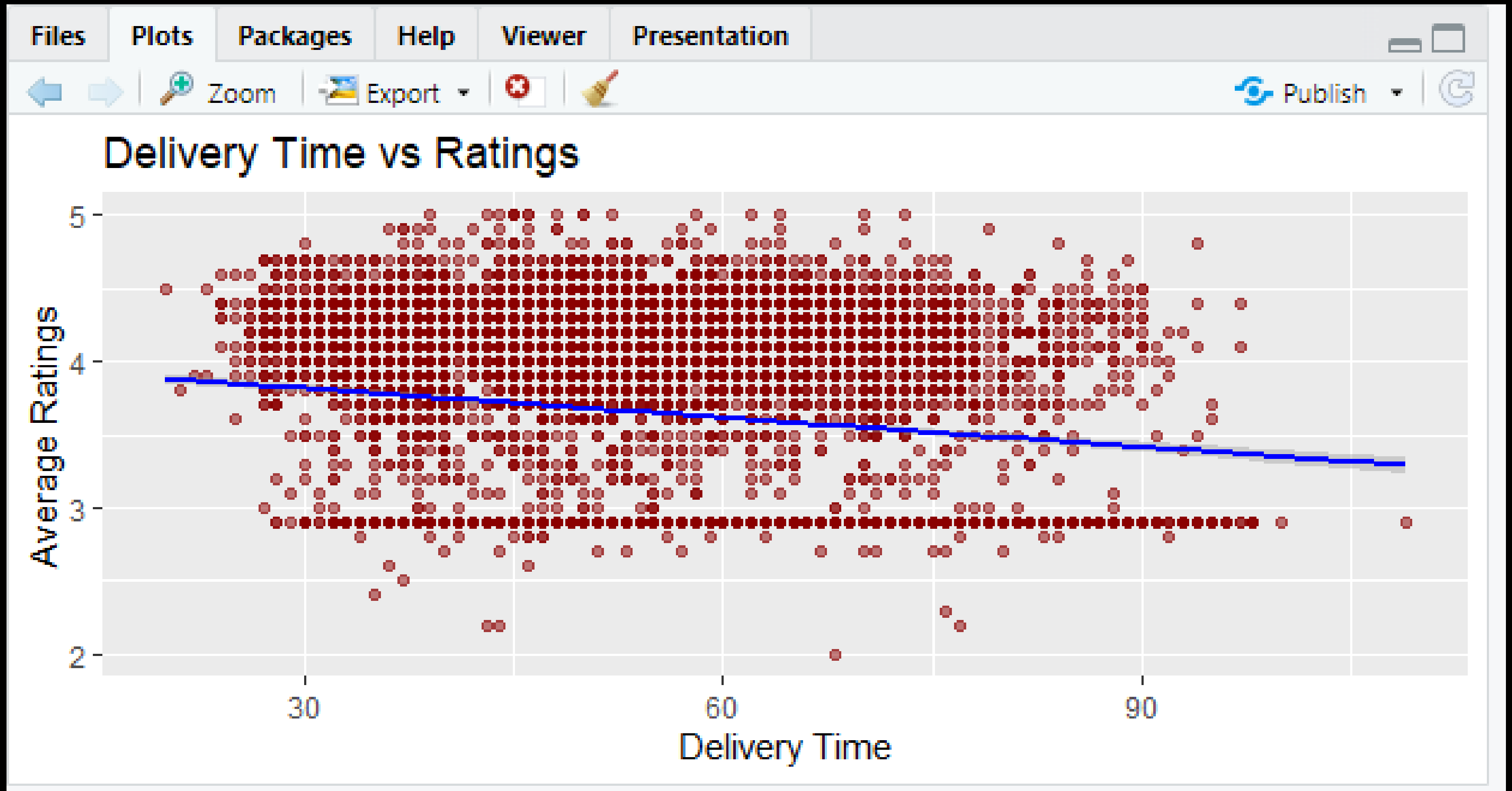


- Key Observations: Hyderabad appears to have the highest total number of reviews, followed by Kolkata. Surat has the lowest number of reviews.
- Insights: This chart provides a quick visual comparison of review activity across different locations. It could be used to identify cities with high levels of user engagement and potentially target marketing or promotional efforts accordingly.

ANALYSIS-15

15. What is the correlation between delivery time and ratings?

- Key Observations: There appears to be a weak negative correlation between delivery time and average ratings. As the delivery time increases, there is a slight tendency for the average rating to decrease. However, the relationship is not very strong, and there's a lot of variability in ratings across different delivery times.
- Insights: The plot suggests that while faster delivery times might be associated with slightly higher ratings, other factors likely play a more significant role in determining customer satisfaction and ratings.



ANALYSIS-16

- **16. What is the average price of meals by area?**

```
Area_price<-swiggy_data %>%
 group_by(Area) %>%
 summarise(Average_Price = mean(Price)) %>%
 arrange(desc(Average_Price))
View(Area_price)
```

Gradedassignment_Bhavyanshu.R* | Area_price

Filter

	Area	Average_Price
1	Vile Parle	1500.0000
2	Pacific Mall Subhash Nagar	1200.0000
3	Someshwarpura	1200.0000
4	Brigade Road	1100.0000
5	Central Market Punjabi Bagh	1050.0000
6	Kurla - Mumbai	1000.0000
7	Mantri Sqare Mall	900.0000
8	Central Markt Punjabi Bagh	875.0000
9	Bandra Area	841.9355
10	Bentinck St.	800.0000
11	Kotturpuram	800.0000

ANALYSIS-17

- 17. What is the total price of all orders in each city?

```
swiggy_data %>%
  group_by(City) %>%
  summarise(Total.Price =
sum(Price))
```

Console   Terminal ×   Background Jobs ×

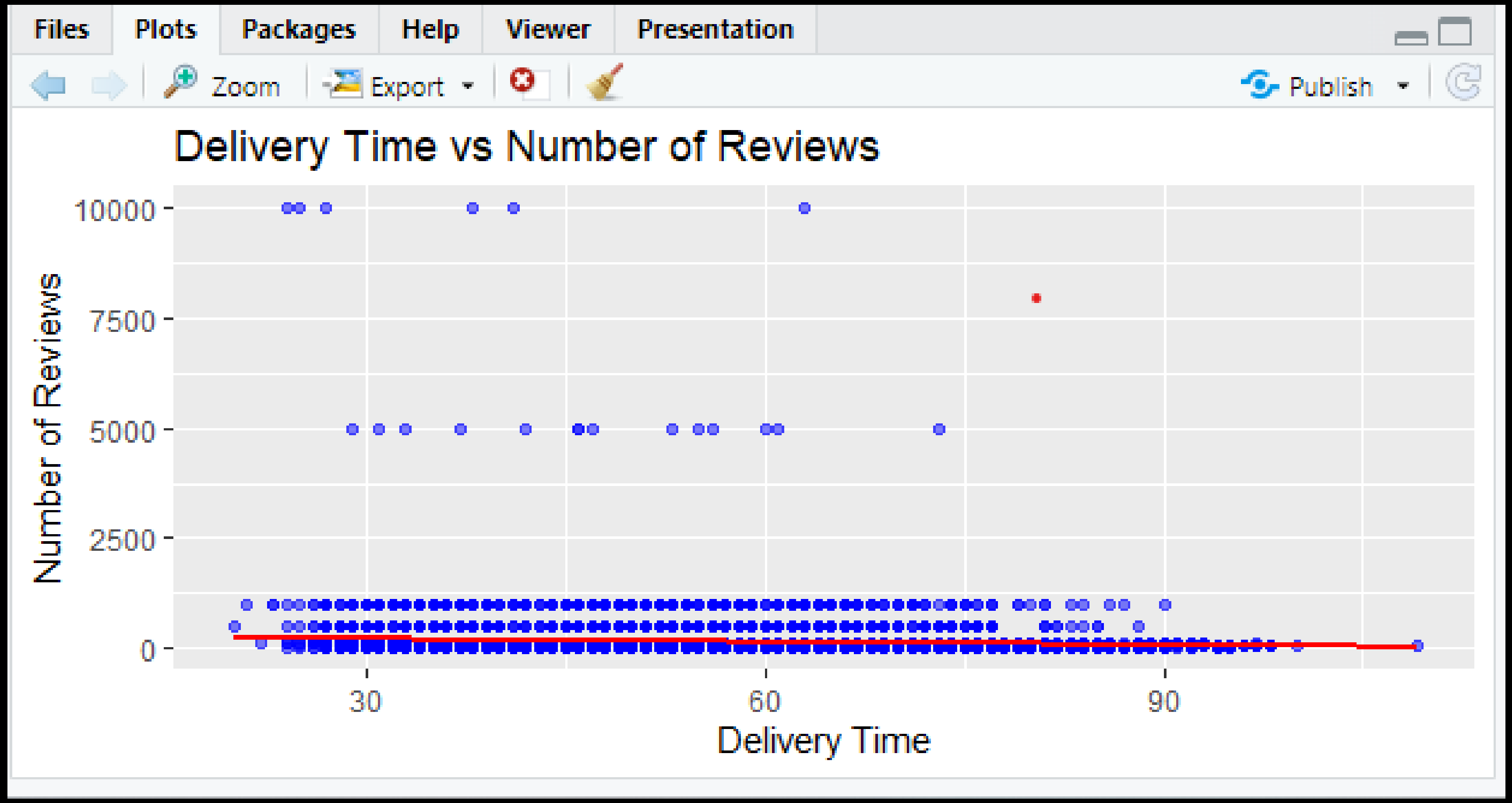
R 4.4.2 · C:/Users/bhavy/AppData/Local/Temp/22

```
# A tibble: 9 × 2
  City      Total.Price
  <chr>           <dbl>
1 Ahmedabad      228098
2 Bangalore      361868
3 Chennai        394010
4 Delhi          203647
5 Hyderabad      322421
6 Kolkata        487648
7 Mumbai         502876
8 Pune           385602
9 Surat          138328
> 
```

ANALYSIS-18

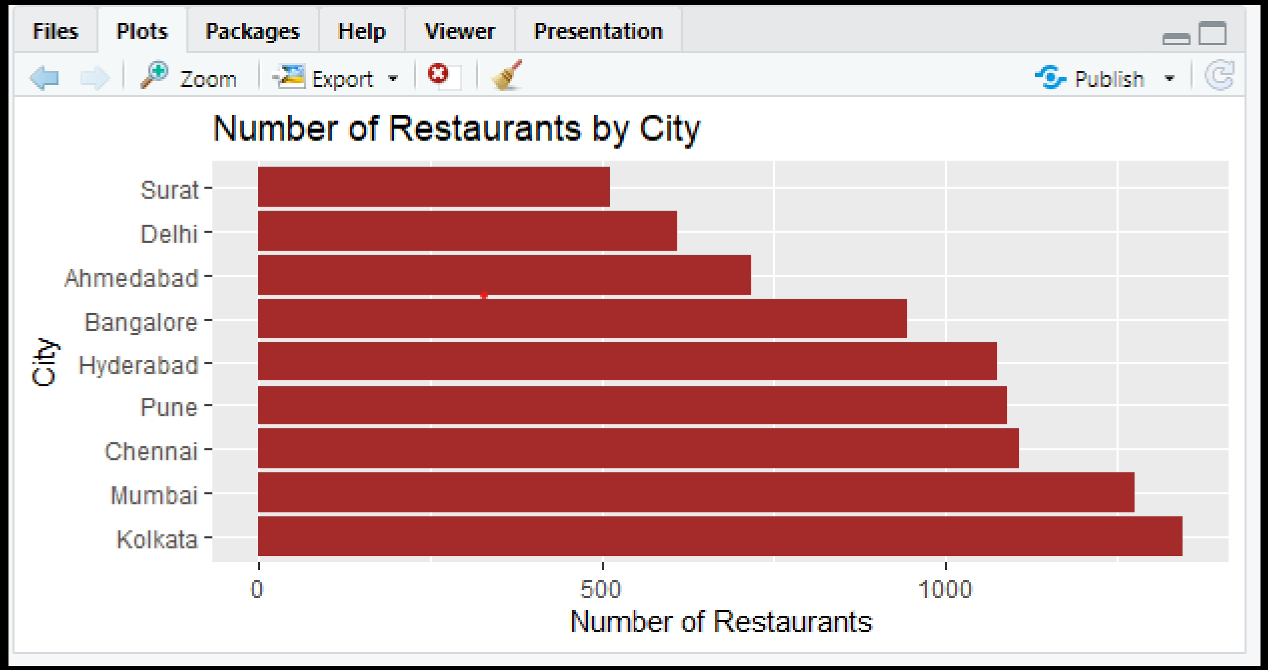
18. Visualize the correlation between delivery time and the number of reviews.

- Key Observations: There appears to be a very weak or negligible relationship between delivery time and the number of reviews. The points are scattered randomly with no clear trend.
- Insights: The plot suggests that delivery time does not seem to have a significant impact on the number of reviews received. Other factors, such as product quality, customer service, or overall popularity, are likely more influential in determining the number of reviews.



ANALYSIS-19

- **19. Which city has the highest number of restaurants? #20. Visualize the number of restaurants by city.**



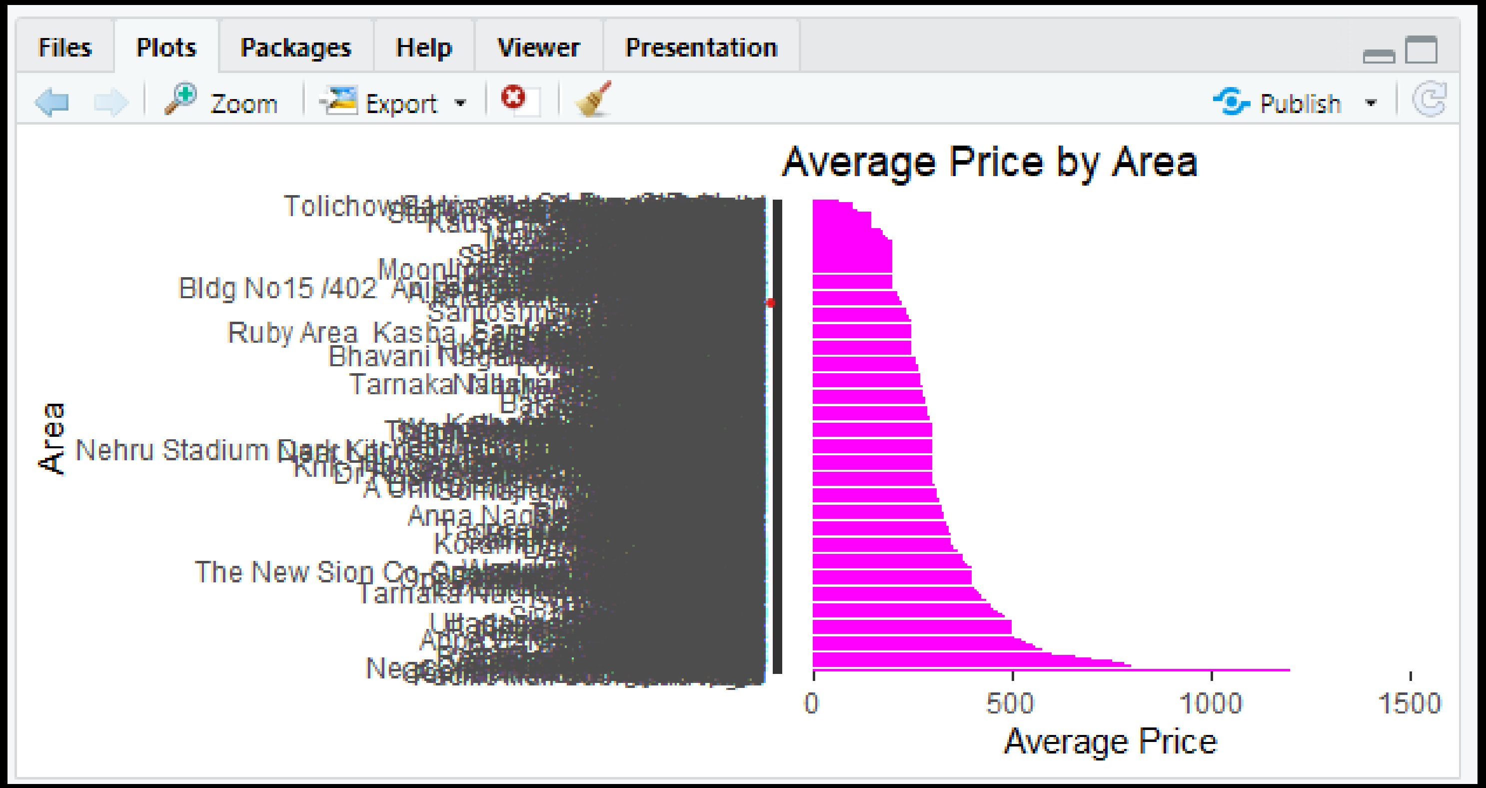

```
swiggy_data %>%
  count(City, name = "Restaurant_Count") %>%
  arrange(desc(Restaurant_Count))

swiggy_data %>%
  count(City, name = "Restaurant_Count") %>%
  ggplot(aes(x = reorder(City, -Restaurant_Count), y =
Restaurant_Count)) +
  geom_bar(stat = "identity", fill = "brown") +
  coord_flip() +
  labs(title = "Number of Restaurants by City", x = "City",
y = "Number of Restaurants")
```

- Key Observations: Kolkata has the highest number of restaurants, followed by Mumbai and Chennai. Surat has the lowest number of restaurants.
- Insights: This chart provides a quick visual comparison of the restaurant density across different locations. It could be used to identify cities with a high concentration of restaurants, which might be relevant for various purposes such as market research, business planning, or urban planning.

ANALYSIS-20

- **20 Visualize the average price of meals by area.**

```
swiggy_data %>%
  group_by(Area) %>%
  summarise(Average_Price = mean(Price)) %>%
  ggplot(aes(x = reorder(Area, -Average_Price), y =
Average_Price)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(title = "Average Price by Area", x = "Area", y =
"Average Price")
```



- Key Observations: There is significant variation in average prices across the areas. Some areas have very high average prices, while others have much lower average prices.
- Insights: This chart provides a quick visual comparison of price levels across different areas. It could be used to identify areas with higher or lower average prices, which might be relevant for various purposes such as market research, business planning, or consumer decision-making.

CONCLUSION

- The analysis highlighted key sales trends, customer preferences, and operational efficiencies.
- Snack Foods and Tier 3 outlets lead sales, while low-fat items and high ratings show a positive correlation with demand.
- Recommendations include optimizing inventory, targeting high-demand products, and refining delivery operations.
- Continuous data monitoring is essential for adapting to market dynamics and ensuring sustained growth.

THANK YOU

2024 Data Analysis Presentation

Bhavyanshu Jain