Data Analysis Report

1. Project Objective/Proposal

The primary objective of this project is to analyze the comprehensive dataset from Doordash to gain insights into its operations spanning across Dashers (drivers), Merchants (restaurants), and Customers. Through this analysis, we aim to understand the current state of the business, identify areas of improvement, and propose actionable recommendations that can enhance the overall efficiency and user experience.

Initial doubts: It is unclear what Placed order with restaurant datetime: Time that restaurant received order is. Because there is a difference between customer placing the order and restaurant receiving the order. Should not this be real time. The definition could have been more precise.

2. Data Cleaning, Feature Engineering, and exploration:

A. Cleaned rows for comparing region wise data along with Time related columns
    1) Null values.
        - Placed order with restaurant datetime
        - Driver at restaurant datetime
    2) Delivery Region='None' for easy comparison between regions.
B. Changed Columns Dtype from object to datetime64[ns]
C. Changed UTC to PT
D. Added Order_ID column for specific data related tracking
E. Changed column names to make it easier to analyze
    1) Customer placed order datetime: Customer_order_time
    2) Placed order with restaurant datetime: Restaurant_order_time
    3) Driver at restaurant datetime: Driver_arrival_time
    4) Delivered to consumer datetime: Delivery_time
F. Added 4 columns that indicates time taken from one stage to another
    1) Cust_to_rest
    2) Rest_to_driv
    3) driv_to_deli
    4) Total_time
G. Deleted 149 rows with negative Time difference, which indicated inconsistency in data entry

Initial Data table details are:

```
Doordash_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18078 entries, 0 to 18077
Data columns (total 13 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Customer placed order datetime        18078 non-null  object
 1   Placed order with restaurant datetime  18038 non-null  object
 2   Driver at restaurant datetime         13547 non-null  object
 3   Delivered to consumer datetime        18078 non-null  object
 4   Driver ID                             18078 non-null  int64
 5   Restaurant ID                         18078 non-null  int64
 6   Consumer ID                           18078 non-null  int64
 7   Delivery Region                       18078 non-null  object
 8   Is ASAP                               18078 non-null  bool
 9   Order total                           18078 non-null  float64
 10  Amount of discount                    18078 non-null  float64
 11  Amount of tip                         18078 non-null  float64
 12  Refunded amount                       18078 non-null  float64
dtypes: bool(1), float64(4), int64(3), object(5)
memory usage: 1.7+ MB
```

After Initial analysis the Table details are:

```
Doordash_df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 13524 entries, 0 to 18077
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Order_ID              13524 non-null  int64
 1   Customer_order_time   13524 non-null  datetime64[ns]
 2   Restaurant_order_time 13524 non-null  datetime64[ns]
 3   Driver_arrival_time   13524 non-null  datetime64[ns]
 4   Delivery_time         13524 non-null  datetime64[ns]
 5   Driver ID             13524 non-null  int64
 6   Restaurant ID         13524 non-null  int64
 7   Consumer ID           13524 non-null  int64
 8   Delivery Region       13524 non-null  object
 9   Is ASAP               13524 non-null  bool
 10  Order total           13524 non-null  float64
 11  Amount of discount    13524 non-null  float64
 12  Amount of tip         13524 non-null  float64
 13  Refunded amount       13524 non-null  float64
 14  Cust_to_rest          13524 non-null  float64
 15  Rest_to_driv          13524 non-null  float64
 16  driv_to_deli          13524 non-null  float64
 17  Total_time            13524 non-null  float64
dtypes: bool(1), datetime64[ns](4), float64(8), int64(4), object(1)
memory usage: 1.9+ MB
```

Insights:

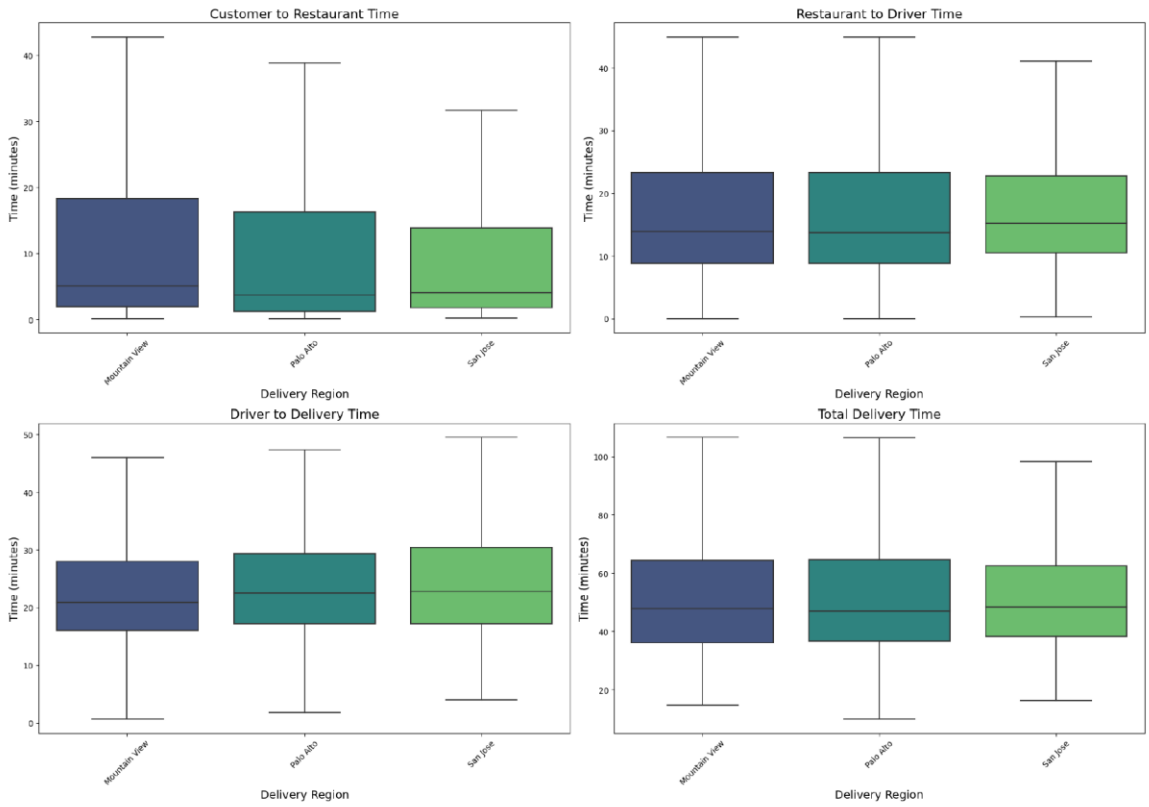
1. Inconsistencies in operation flow as 149 times the usual flow of ordering i.e. customer, restaurant, driver, and delivery is not followed.
2. “Driver at Restaurant datetime" is null 4531 times out of 18077 entries. (25%).
   Questions that Arise:
   a. Is this a manual task performed by driver/restaurant?
   b. Can this be a software glitch?
   c. Does the Driver/restaurant forget to enter driver's arrival at restaurant because of time pressure?
   d. What is the reason for null values.

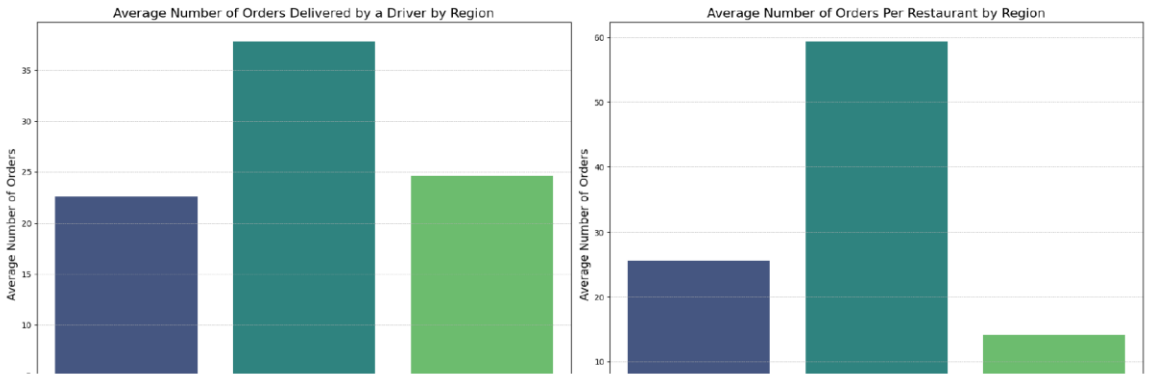
Recommendation:

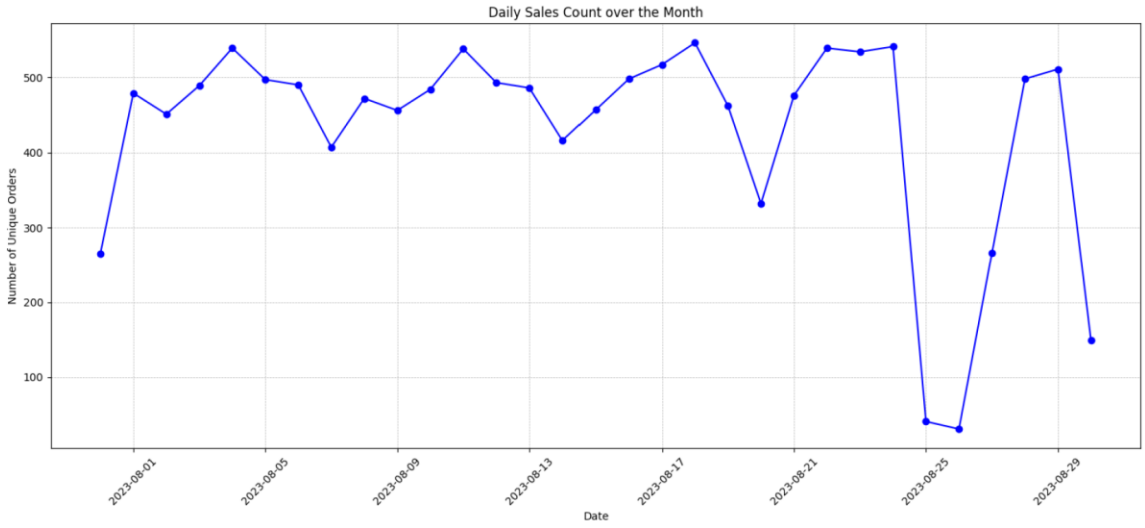
DoorDash needs to look at the data entry issue on column “Driver at Restaurant datetime” to analyze operational efficiency more effectively to provide faster deliveries and more insights.

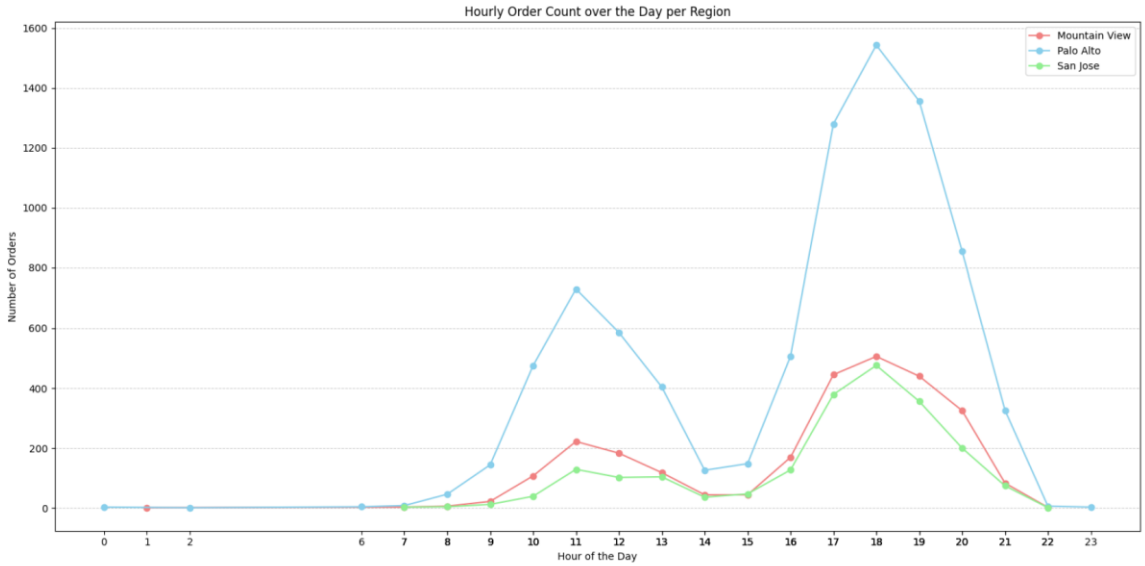
Data Analysis & Recommendation

Delivery Time Analysis:









1. Palo Alto has the least median Total delivery time, customer to restaurant delivery time, restaurant to driver time. Mountain view tops in driver to delivery time. Looks like Sanjose has

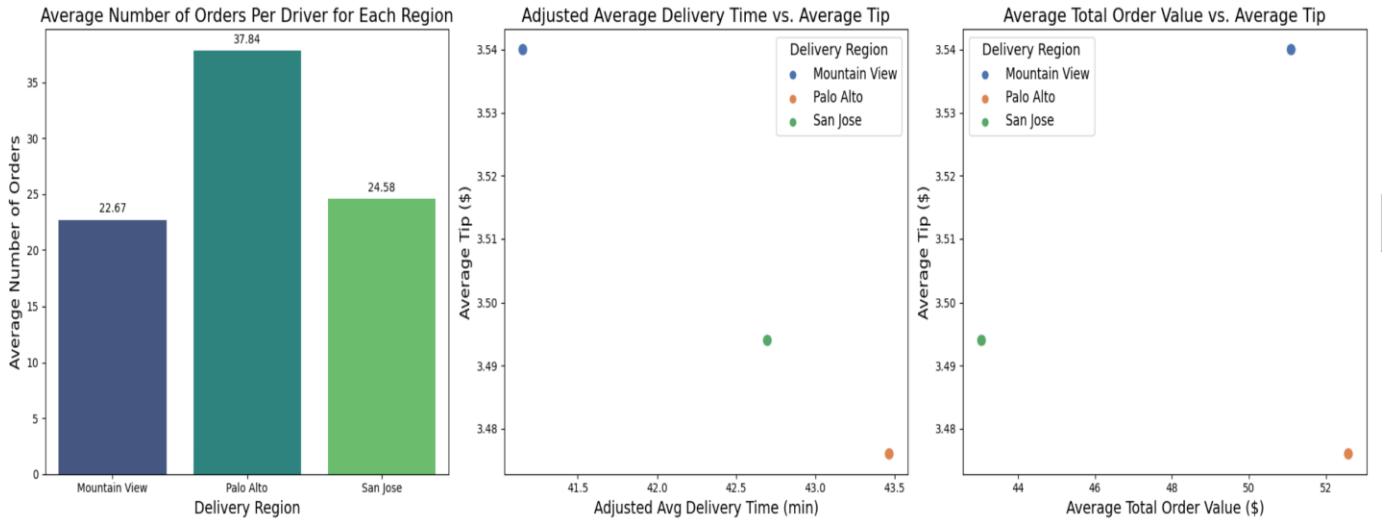
the highest total delivery time, which means less customer satisfaction because of longer waiting times.

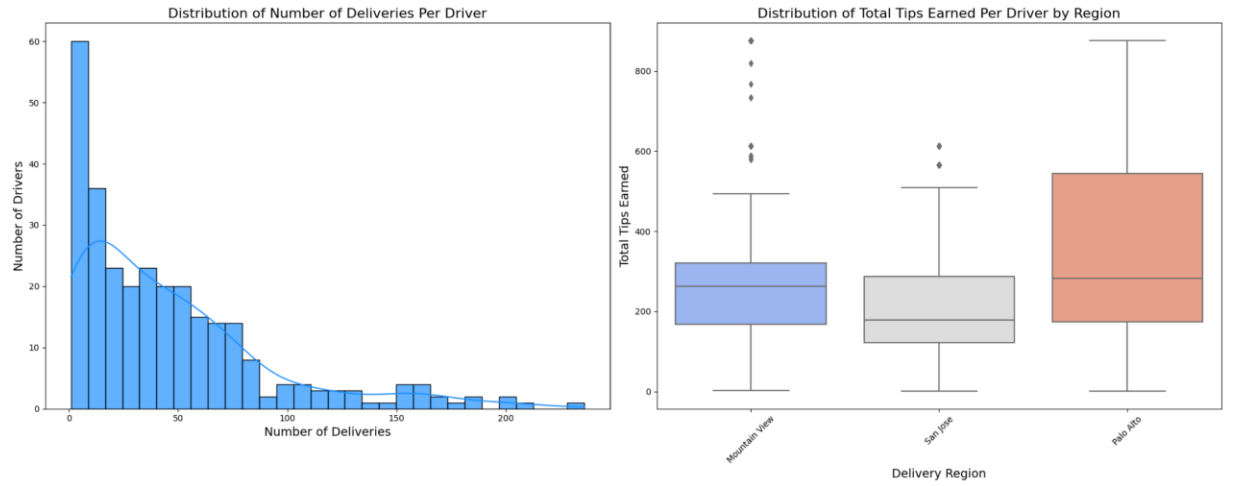
2. Interestingly, the Merchants in Palo Alto sell twice the number of orders compared to other two region, but still, they have least time from customer to restaurant and restaurant to driver. This clearly indicates their high performance in managing multiple orders.

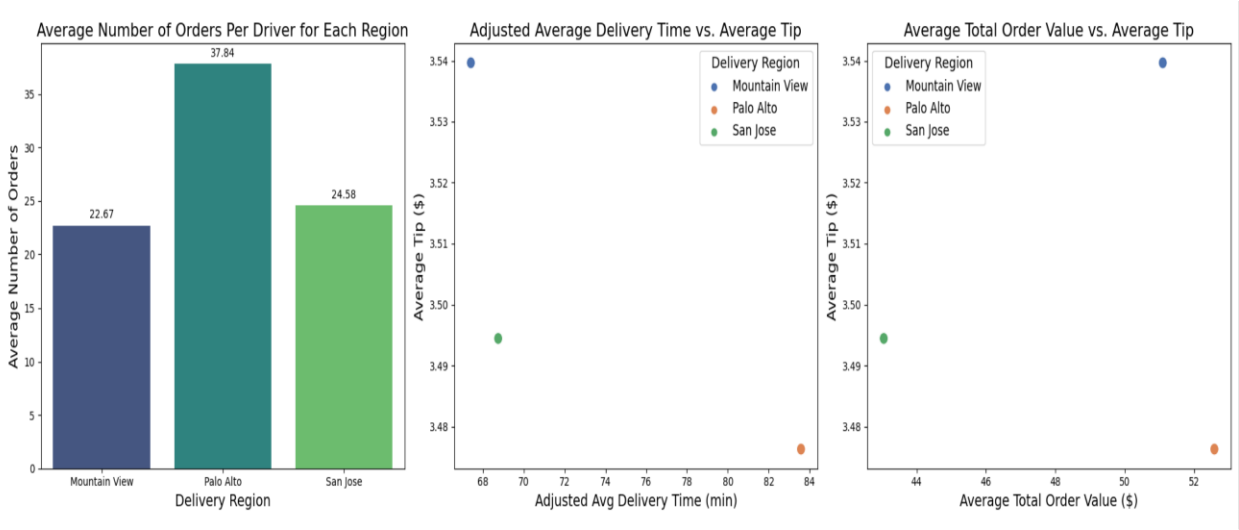
Recommendation

1. Merchant Sales improvement: It would be helpful for Doordash to compare the operations of Merchants by region and apply some learnings from Palo Alto to others. This step can help merchants in Mountain view and San Jose to deliver order faster, gain more orders and revenue, with increased operational efficiencies per order.
2. Customer Delight with faster deliveries.
3. Merchant workplace optimization:
    a. The Peak hours of the day based on sales can be analyzed by Merchants to improve operations and cut costs.
    b. The Daily sales count is low on 25th and 26th of the month. If we have yearly data, Merchants can be given insights to plan leaves for their employees.

Dashers:







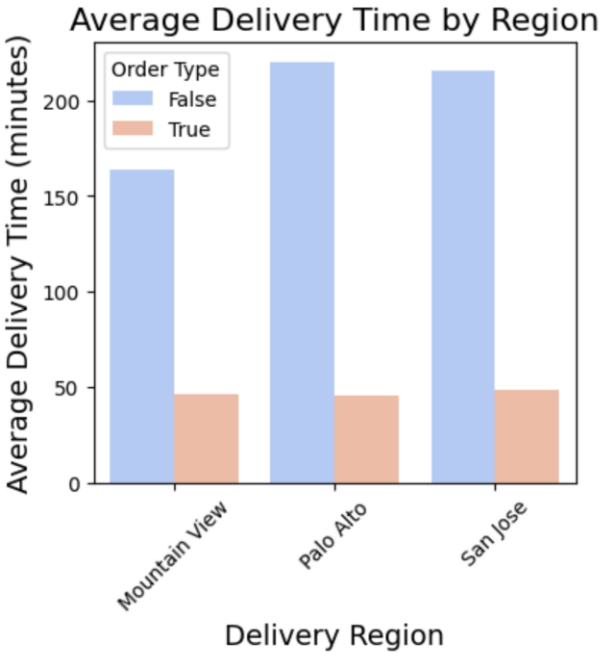
1. Average number of orders per driver is highest for Palo Alto, indicating they are either full time or they work long hours.
2. The average tip amount is least for Dashers at Palo Alto, though their average delivery time and average order values are high compared to dashers in other regions.
3. Interestingly, total tip earned by a dasher in palo alto is more as they deliver on average more orders.

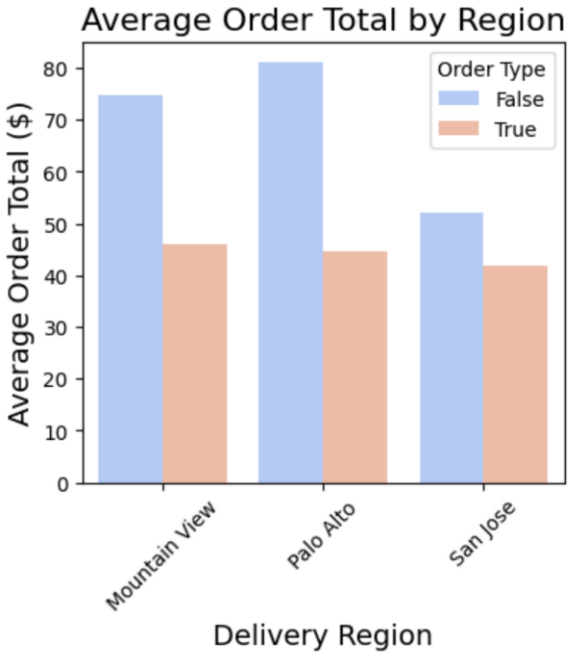
Recommendation

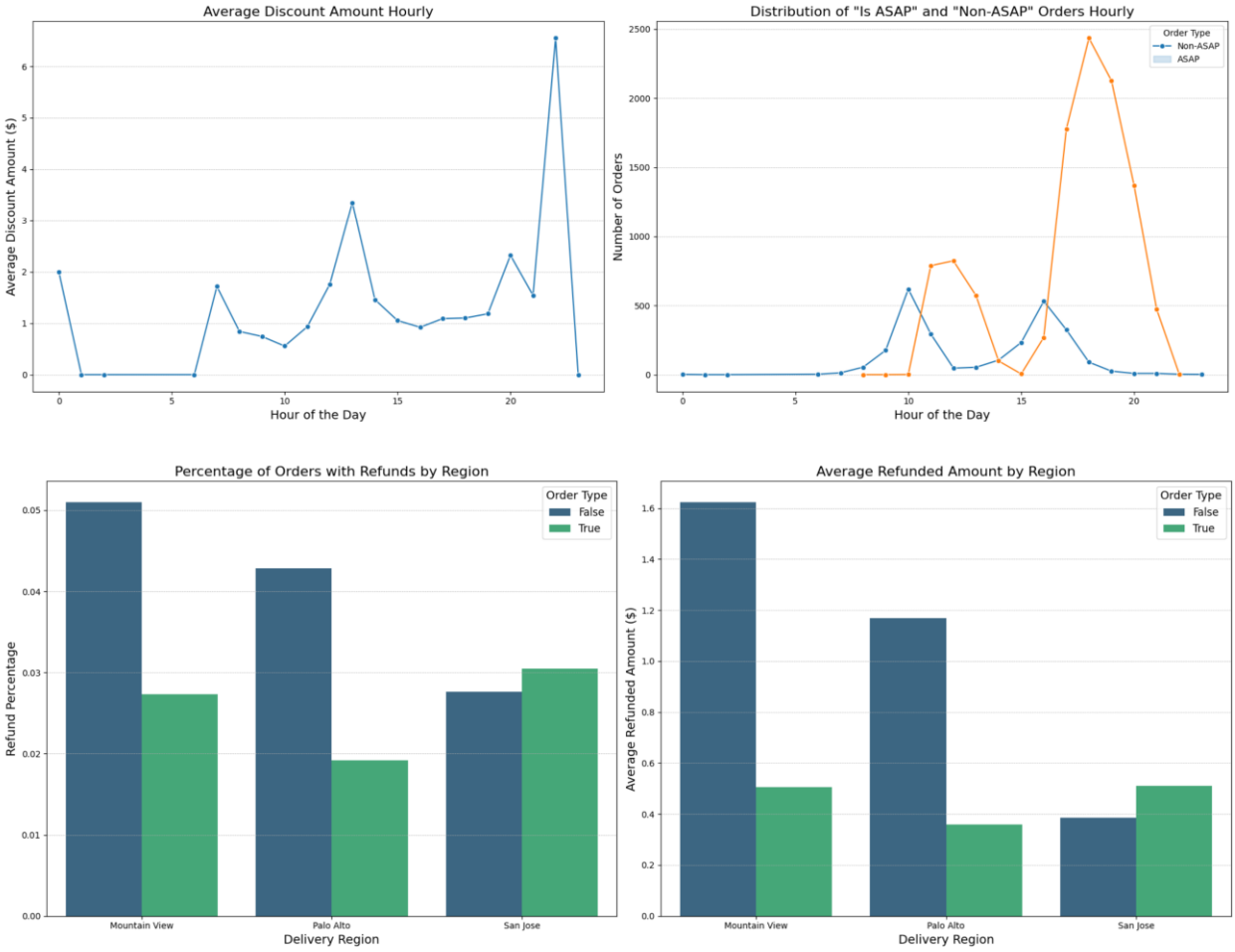
1. “No one likes to live in uncertainty, unfortunately dasher tips are uncertain, but DoorDash can help them”
2. Dasher income improvement program: Many Dashers consider Tip as a major source of income in addition to salary. The company can show them the data of average tips and help them understand their revenue potential. This will motivate the Dasher to complete more orders per day by increasing his performance and his speed of delivery. This also helps them plan their schedules effectively.

3. Dasher bonuses: DoorDash might want to think of a plan if certain Dashers are consistently not paid well through tips.

Non-ASAP over ASAP:







1. Average Order Total and Average Tip amounts are higher, indicating merchants can leverage this to improve revenue. Also, this indicates Customers are happy with Non-ASAP and tend to give more tip. This can be because of discounts for Non ASAP, but we need further analysis for this trend.
2. Looks like the average refunded amount is higher, which corresponds to average total order, so that is relatable.
3. Customers are delighted by discounts and have larger order size Non-ASAP orders, indicating customers want such discounts.
4. For Non-ASAP the refund% is high for Mountain view, and for ASAP the refund% is high for San Jose.

Recommendation

1. Discount Notification: It might be helpful to send notifications or show these discount times to customers to save on purchase or order more. This brings customer delight.
2. Merchant Revenue vs profit optimization: Merchants can also be enticed to give discounts to receive more orders by showing them the analytics for their stores. Doordash can help them with breakeven analysis for discounts.
3. Zero Refund Literacy program: DoorDash should analyze the cause of refund seriously, as that is loss for customer, merchant, and DoorDash. Doordash should conduct regional data comparison

along with sentiment analysis. Then train the Merchants and dashers to ensure customer satisfaction.