Data Report

This report presents a project overview for the DoorDash DashMart - Vendor Analytics role, delving into the delivery system to uncover patterns, trends, and potential opportunities. Each section offers actionable insights, ranging from dasher efficiency and store performance assessment to customer satisfaction and cancellation trend analysis, all of which could significantly impact the business landscape.

My approach to this exploratory data analysis (EDA) began by converting the provided Excel sheet into a .csv file and storing it locally. Utilizing Python within VS Code, I employed a Jupyter Notebook extension for data manipulation and analysis. Accompanying this report is a Python file containing the EDA process and its associated visualizations.

Commencing with the import of essential libraries (pandas, numpy, matplotlib, etc.) and the dataset itself, I conducted an initial data assessment using functions like data.describe() and data.info(). Null value counts were evaluated, and a reference database was established (Refer to Appendix - I (a)).

1) Data Cleaning and Exploration

Continuing with the analysis, the next step involved assessing the data's cleanliness and structure. The dataset comprised 60,583 rows and 19 columns, demonstrating its substantial scope.
I conducted a thorough review of each feature's minimum and maximum values to identify potential outliers, ensuring their adherence to acceptable ranges. Detailed records can be found in Appendix - I(b).

The further examination involved grouping by delivery ID to determine the item count within each delivery. To enhance data organization, the DELIV_CREATED_AT column was divided into separate DATE and TIME columns. The original DELIV_CREATED_AT column was subsequently dropped, and the modified dataset was stored separately.

Reformatting the dataset for clarity, an array of new column names was generated, positioning DATE and TIME as the second and third columns.

With data cleanliness verified and necessary modifications made, I proceeded to initiate the Exploratory Data Analysis (EDA) phase.

2) Exploratory Data Analysis

The exploration of the dataset commenced with the objective of identifying notable trends and patterns, specifically focusing on conducting time series analyses. The following analyses were undertaken:

a) Analysis of Order Frequency by Hour (Refer to Appendix - I(c))

To facilitate this analysis, the Date and Time features were converted to the datetime data type using the "pd.to_datetime" function.

From the resulting plot, the following insights were extracted:

- Prominent hours of order placement became evident, notably during lunch and dinner periods (11, 12, 17, 18, 19, 20, 21). These spikes suggest peak order times, providing valuable guidance for the allocation of delivery resources. Enhancing dasher staffing during these hours could enhance efficiency in handling increased demand.

- Optimizing Peak Hours:

Action: Implement dynamic dasher scheduling to accommodate peak hours effectively.

Recommendation: Utilize real-time data to predict and manage surges in orders. Optimize dasher assignments during high-demand periods.

- Promotions During Off-Peak Hours:

Action: Focus on generating sales during low-activity periods.

Recommendation: Drive orders during off-peak hours by offering targeted promotions, discounts, or loyalty rewards.

- Sustainable Growth Strategy:

Action: Align business expansion with demand patterns.

Recommendation: Develop a growth strategy that balances expansion goals with service quality, based on historical data.

It is advised to collaborate with various teams, including marketing, business development, operations, and technology, to execute the recommendations effectively and realize their benefits.

b) Analysis of Dasher Deliveries per Month (Refer to Appendix - I(d))

During the months of September and October, a comprehensive analysis was conducted to determine the number of deliveries completed by each driver. The data indicated a total of 3,033 dashers actively engaged during this period.

deliveries_per_driver

	DELIV_DASHER_ID	year	month	DELIVERY_UUID
0	217256.0	2022	10	6
1	296959.0	2022	10	2
2	520620.0	2022	10	5
3	606237.0	2022	9	37
4	606237.0	2022	10	122
...	...	...	...	...
3028	22864408.0	2022	10	5
3029	22878812.0	2022	10	4
3030	22888318.0	2022	10	3
3031	22891158.0	2022	10	2
3032	22919415.0	2022	10	3

3033 rows × 4 columns

c) Identification of Top Performing Dashers per Month (Refer to Appendix - I(e))

The analysis focused on identifying the top-performing dashers within each month. Noteworthy insights and corresponding recommendations emerged from the analysis, contributing to enhanced delivery efficiency:

- Recognition of High Performers:

Insight: Certain dashers consistently excel by delivering a high number of orders monthly.

Recommendation: Implement a recognition and incentive system to acknowledge the efforts of top dashers, which can include bonuses, rewards, or priority scheduling.

Execution Approach: Establish a structured rewards program that evaluates and commends dashers based on their monthly performance. Collaboration with the Human Resources or Operations team is recommended to design and execute effective incentive strategies.

- Diversification of Dasher Pool:

Insight: Over-reliance on a limited group of top dashers carries inherent risks.

Recommendation: Foster a diverse and broad spectrum of dasher participation, mitigating potential disruptions caused by the unavailability of top performers.

Execution Approach: Actively promote the recruitment of dashers from varied demographics and geographical locations. Utilize data insights to pinpoint untapped demographics and strategically target them for recruitment.

Implementing these recommendations has the potential to optimize dasher management, ensuring both efficient operations and consistent service quality.

d) Calculation of Order-to-Delivery Time:

The attempt to analyze the time elapsed between order placement by the customer and subsequent delivery encountered challenges due to timestamp formatting discrepancies.

e) Analysis of Daily Delivery Counts (Refer to Appendix I - f)

The following insights emerged from the examination of daily delivery counts, yielding recommendations for operational enhancements:

- Optimization of Peak Day Operations:

Insight: Weekends witness heightened delivery activity, potentially straining operations.

Recommendation: Formulate a dedicated operational strategy tailored to high-traffic days (weekends) to ensure streamlined execution.

Execution Approach: Develop a comprehensive plan encompassing optimized dasher scheduling, inventory management, and customer communication strategies for weekends. Incorporate dynamic resource allocation to promptly adjust staffing according to real-time demand.

- Response to Seasonal Fluctuations:

Insight: Daily delivery volumes exhibit fluctuations, with certain days displaying significantly higher delivery counts. Such variance may stem from factors like holidays, promotions, or events.

Recommendation: Thoroughly investigate the drivers behind these fluctuations, pinpointing the influence of holidays or events. Craft targeted marketing initiatives and special promotions for these occasions.

Execution Approach: Harness historical data analysis to identify trends during holidays and events. Create tailored promotions or discounts suited to these periods. Leverage data analytics to predict demand and optimize inventory management.

- Effective Supply Chain and Inventory Management:

Insight: Elevated delivery volumes on specific days can exert additional pressure on supply chain and inventory management. Ensuring ample stock availability to meet demand remains pivotal.

Recommendation: Implement predictive inventory management solutions that factor in historical demand patterns to optimize stock levels and avert stock outs during peak periods.

- Analysis of Customer Feedback:

Insight: Scrutinize customer feedback and ratings during high-volume delivery days to discern areas necessitating improvement.

Recommendation: Utilize customer feedback to address pain points and elevate the overall delivery experience during peak times. This effort can foster heightened customer satisfaction and loyalty.

Implementation of these recommendations promises to elevate operational efficiency, customer satisfaction, and overall service quality during high-demand intervals.

f) Analysis of Deliveries per Store (Refer to Appendix I - g)

The examination of deliveries per store yielded the following insights, with subsequent recommendations for optimization:

- Store-Specific Order Distribution:

Insight: DashMart dominates the order count, signifying its high popularity.

Recommendation: Delve deeper into DashMart's performance, particularly by exploring its product categories.

Execution Path: Analyze DashMart's internal performance metrics and its products' sales data. Identify top-performing categories and strategize ways to maintain and expand their success.

g) Analysis of Deliveries per Individual Customer (Refer to Appendix I - h)

The analysis of individual customer delivery frequencies led to the following insights and corresponding recommendations:

- Customer Engagement Stratification:

Insight: Diverse customer engagement levels emerge based on order frequency.

Recommendation: Implement a tiered engagement system correlating with delivery frequency (e.g., Bronze, Silver, and Gold tiers), each offering increasing benefits.

Execution Path: Craft a loyalty program where customers ascend tiers based on order frequency. Introduce rewards such as discounts, early product access, and exclusive support for higher tiers.

- Resuscitation Initiatives:

Insight: Low-activity customers warrant incentives for increased ordering.

Recommendation: Launch personalized re-engagement campaigns targeting such customers.

Execution Path: Deploy tailored email campaigns featuring exclusive discounts or offers to stimulate additional orders. Tailor re-engagement strategies according to distinct customer segments.

- Subscription Proposition:

Insight: Subscriptions may resonate with some customers seeking convenience.

Recommendation: Introduce subscription models for frequently purchased items.

Execution Path: Develop subscription packages featuring convenience and cost savings. Allow customers to select the delivery frequency and items. Promote subscriptions through targeted marketing, concentrating on medium to high-order frequency customers.

- Unique Product Access:

Insight: High-frequency customers display a strong affinity for both products and services.

Recommendation: Extend exclusive access to novel products or time-limited offerings.

Execution Path: Dispatch early-access notifications to high-frequency customers, enticing them with distinct products or discounts before wider availability. Utilize personalized recommendations based on order history.

- Collaborative Endeavors:

Insight: Collaborations can allure diverse customer segments.

Recommendation: Forge partnerships with local businesses for joint promotions.

Execution Path: Identify businesses that align with your brand identity and attract a comparable customer base. Create collaborative offers like bundle deals or cross-promotions, enticing customers from both entities.

h) Examination of Order Substitutions (Refer to Appendix I - i)

The investigation into order substitutions produced the subsequent insights and recommendations:

Substitution Occurrence: Approximately 3.48% of orders involve substitutions.

Alignment Rate: Among substituted items, only 2.86% align with customers' original preferences.

Recommendations:

- Substitution Precision Enhancement: Given the low alignment rate, prioritize enhancing substitution accuracy. Select alternatives closely resembling the original items to foster customer contentment, trust, and loyalty.

Short Term (Immediate Action): Develop a database for frequently requested substitutions, aiding dashers in fulfilling orders more accurately.

Long Term (Strategic Approach): Leverage machine learning algorithms to predict suitable substitutions. Utilize historical data to train models, subsequently refining them over time.

- Algorithm Refinement: Invest in optimizing the algorithms suggesting substitutions. Employ customer data, preferences, and historical choices to propose alternatives that customers are likely to accept. Implement machine learning models for learning from prior substitution patterns.

- Customer Feedback Integration: Institute a feedback loop involving customers regarding substitutions. Allow customers to voice their thoughts on substitutions received. Utilize this feedback to consistently enhance the quality of substitutions, tailoring them to preferences.

Short Term (Immediate Action): Encourage customers to provide feedback through post-order surveys.

Long Term (Strategic Approach): Incorporate an integrated feedback mechanism within the app. Incentivize customers to provide feedback leading to improved substitution accuracy.

- Diversity in Substitute Options: Expand the assortment of substitute items accessible to dashers. Provide a broader array of alternatives for common items, thereby heightening the likelihood of finding suitable substitutes aligned with customer preferences.

Short Term (Immediate Action): Regularly update the list of available substitutes based on popularity and dasher recommendations.

Long Term (Strategic Approach): Utilize customer demographic data to offer a diverse array of substitute items catering to varied tastes and dietary preferences.

Applying these recommendations can foster a more accurate, satisfying, and efficient substitution process, elevating the overall customer experience.

i) Identification of Missing or Incorrect Items (Refer to Appendix I - j)

With 208 instances accounting for a mere 1.59%, the dataset demonstrates a relatively low occurrence of missing or incorrect items. While the percentage is small, any instance of missing or incorrect items can impact customer satisfaction and trust. Therefore, ongoing monitoring and corrective actions are vital.

Upon further investigation, additional analysis can be conducted to determine if factors like time, item, or dasher activity influence the occurrence of missing or incorrect items.

j) Time Analysis, Store and Dasher Analysis, Item Category Analysis (Refer to Appendix I - k) for the Entire Dataset

This comprehensive analysis yielded several insights, highlighting peak dasher activity on Sundays and Fridays, the popularity of "Drinks" on DashMart, and DashMart1's prevalence in dasher visits.

k) Customer Satisfaction Rating (Refer to Appendix I - l)

With an impressive overall customer satisfaction rate of 97.81%, the majority of customers express contentment with their orders. This signifies that the delivery service largely meets expectations concerning order accuracy, item availability, and overall delivery experience.

Recommendations include maintaining consistency, staying attuned to customer feedback, and addressing any exceptions to uphold high customer satisfaction.

l) Analysis of DELIV_CANCELLED_AT for Cancellation Trends (Refer to Appendix I - m)

The 1.57% cancellation rate indicates a low occurrence of order cancellations, generally reflecting customer satisfaction with the service. To further enhance customer experience, variations in cancellation rates based on specific days should be closely scrutinized. Identifying common reasons for cancellations can facilitate focused improvements.

m) Store Performance Analysis and Calculation of Delivery Times and Order Accuracy (Refer to Appendix I - n)

The store performance analysis provides valuable insights into delivery times and order accuracy for different stores.
Recommendations include:
Improving delivery efficiency by optimizing routes and resource allocation.
Enhancing order accuracy through accurate packing and fulfillment.
Leveraging best practices from top-performing stores.
These recommendations aim to elevate operational efficiency, enhance customer satisfaction, and maintain a high standard of service quality.

n) Analysis of Dasher Performance Metrics (Average Delivery Time, Order Missing Accuracy, Order Substitution Accuracy) - Appendix I (n)

Average Delivery Time: Significant variations exist in dasher performance regarding average delivery time. Certain dashers consistently achieve faster delivery times compared to others.
- Order Missing Accuracy: Most dashers demonstrate a notably low order missing accuracy, indicating infrequent instances of missing items from customer orders.

- Order Substitution Accuracy: Some dashers exhibit high order substitution accuracy, suggesting their effective selection of suitable substitutes for unavailable items.

- Delivery Punctuality: The majority of dashers maintain a perfect delivery punctuality score, consistently delivering orders on time.

- Customer Satisfaction: Predominantly high customer satisfaction scores indicate customers' general contentment with dasher service.

Recommendations:

- Training and Best Practices Sharing: Identify and encourage top-performing dashers within each category to share their effective practices with others, elevating overall performance.

- Performance Recognition & Monitoring: Establish a recognition program to spotlight and reward exemplary dashers. Continuous performance monitoring and feedback mechanisms will help dashers gauge their progress and make improvements.

o) Analysis of Price Distribution across Items Using ITEM_PRICE (Appendix I - o)

Average Price: $5.93

Median Price: $4.49

Minimum Price: $0.29

Maximum Price: $149.50

Distribution Pattern: The analysis indicates that more orders are placed for cheaper items compared to expensive ones. The majority of orders involve items priced under $10.

p) Dasher Acceptance Patterns Analysis (CLAT) - Appendix I (p)

Issues regarding time conversions and manipulations affected accurate analysis.

q) Analysis of Correlation between Item Price and Likelihood of Substitutions (Appendix I - q)

Substitution Rate vs. Item Price Range: A noticeable correlation exists between item price and the probability of substitutions. As item price increases, the likelihood of substitutions also rises. This suggests that higher-priced items may warrant special attention to ensure accurate and available substitutions, thereby enhancing customer satisfaction.

r) Time-of-Day Analysis for Delivery Times, On-Time Rates, and Order Accuracy - Appendix I (r)

```
   Time of Day  Average Delivery Time (minutes)  On-Time Rate  Order Accuracy
0    Afternoon                    3.681559e+07      0.038100        0.930856
1      Evening                    3.681533e+07      0.050543        0.926483
2      Morning                    3.681607e+07      0.066667        0.933333
3        Night                    3.681587e+07      0.049129        0.952363
/var/folders/n_/z79s6yq55dqdq9yn_lhckr6m0000gn/T/ipykernel_48324/1812457429.py:14:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

- Average Delivery Time Variation: Delivery times differ based on the time of day. Morning deliveries tend to be quicker, while night time deliveries take longer.

- On-Time Rate: Night-time deliveries show a lower on-time rate, indicating potential challenges like navigation or dasher availability during that time.

- Order Accuracy: Overall order accuracy remains high throughout the day, with a slight dip in the evening. Increased order volume during the evening might lead to potential errors.

Recommendations:

- Enhance Night Deliveries: Address on-time rate challenges for nighttime deliveries through incentives or route optimization to improve customer satisfaction.

- Leverage Morning Efficiency: Capitalize on efficient morning deliveries with potential promotions to encourage more orders during that time.

- Improve Evening Accuracy: Investigate the evening accuracy dip, possibly through dasher training or better inventory management.

- Optimize Scheduling: Utilize these insights for dasher scheduling, allocating resources during peak times and streamlining operations during slower periods.

(a)   Null Database

	column_name	null_count
8	DELIV_CANCELLED_AT	59914
18	SUBSTITUTE_ITEM_CATEGORY	57995
17	SUBSTITUTE_ITEM_NAME	57995
7	DELIV_CLAT	2393
5	DELIV_D2R	1287
3	DELIV_DASHER_ID	1009
12	WAS_SUBBED	0
16	ITEM_CATEGORY	0
15	ITEM_PRICE	0
14	ITEM_NAME	0
13	WAS_FOUND	0
0	DELIVERY_UUID	0
11	WAS_MISSING	0
10	WAS_REQUESTED	0
1	DELIV_CREATED_AT	0
6	DELIV_IS_20_MIN_LATE	0
4	DELIV_SUBMARKET	0
2	DELIV_STORE_NAME	0
9	DELIV_MISSING_INCORRECT_REPORT	0

```
null_df.to_csv('/Users/preethi/Downloads/null_cou
```

(b)   Table for min max values

	DELIV_DASHER_ID	DELIV_D2R	DELIV_IS_20_MIN_LATE	DELIV_CLAT	WAS_REQUESTED	WAS_MISSING	WAS_SUBBED	WAS_FOUND	ITEM_PRICE
min	217256.0	0.0	0.0	0.050000	1.0	0.0	0.0	0.0	0.29
max	22919415.0	1313.5	1.0	119.783333	1.0	1.0	1.0	1.0	160.95

(c) Number of orders placed vs hour of day graph
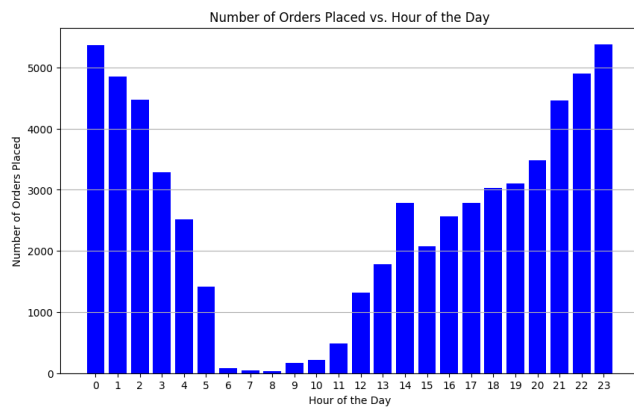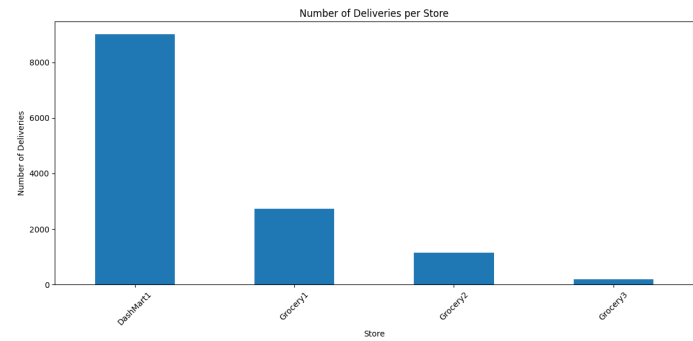

(d) Number of Deliveries per Dasher per Month
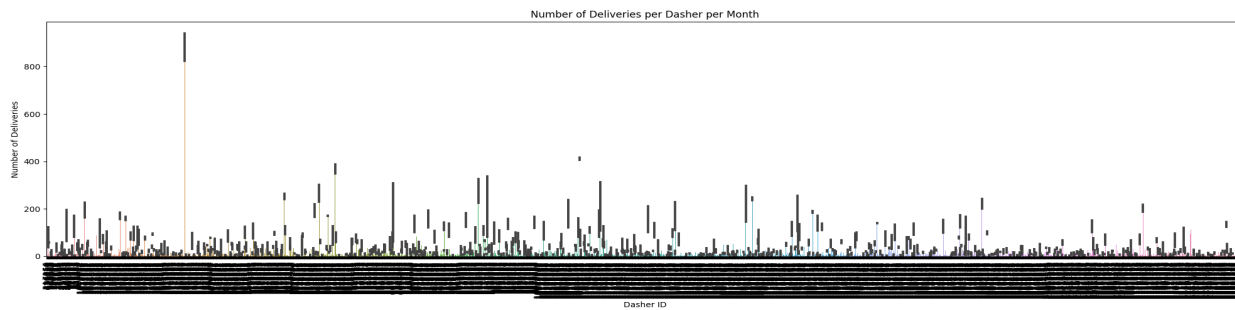
(e) Top 10 performing dashers per month

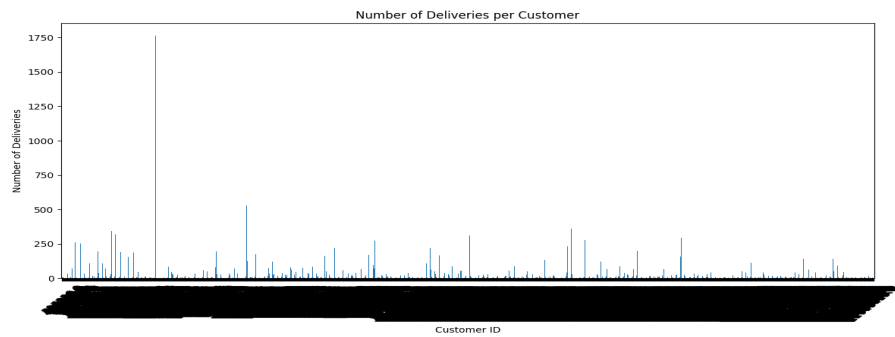|	DELIV_DASHER_ID	year	month	DELIVERY_UUID
10	2446877	2022	10	941
11	10910871	2022	10	406
12	5784382	2022	10	387
13	8841279	2022	10	337
14	11428124	2022	10	313
15	6619837	2022	10	309
16	4821259	2022	10	265
17	18892072	2022	10	243
18	14827828	2022	10	236
19	5456278	2022	10	230
0	2446877	2022	9	824
1	10910871	2022	9	417
2	5784382	2022	9	349
3	8616812	2022	9	326
4	5456278	2022	9	301
5	14622346	2022	9	298
6	15510588	2022	9	255
7	14827828	2022	9	249
8	4821259	2022	9	241
9	10703068	2022	9	237

f) Number of deliveries per day

g)

h)

i)

j) (i)

Time Analysis

(ii)

Store and Dasher Analysis

(iii)

Item Category Analysis

k)

Time Analysis

Top Stores and Dashers

Top Item Categories

l)

Overall Customer Satisfaction

m)

Cancellation Trends

n)

Distribution of Item Prices

o)

Item Price Range vs. Substitution Rate