X Analytics Case

Mayur Bhatia

Summary & Assumptions

Summary

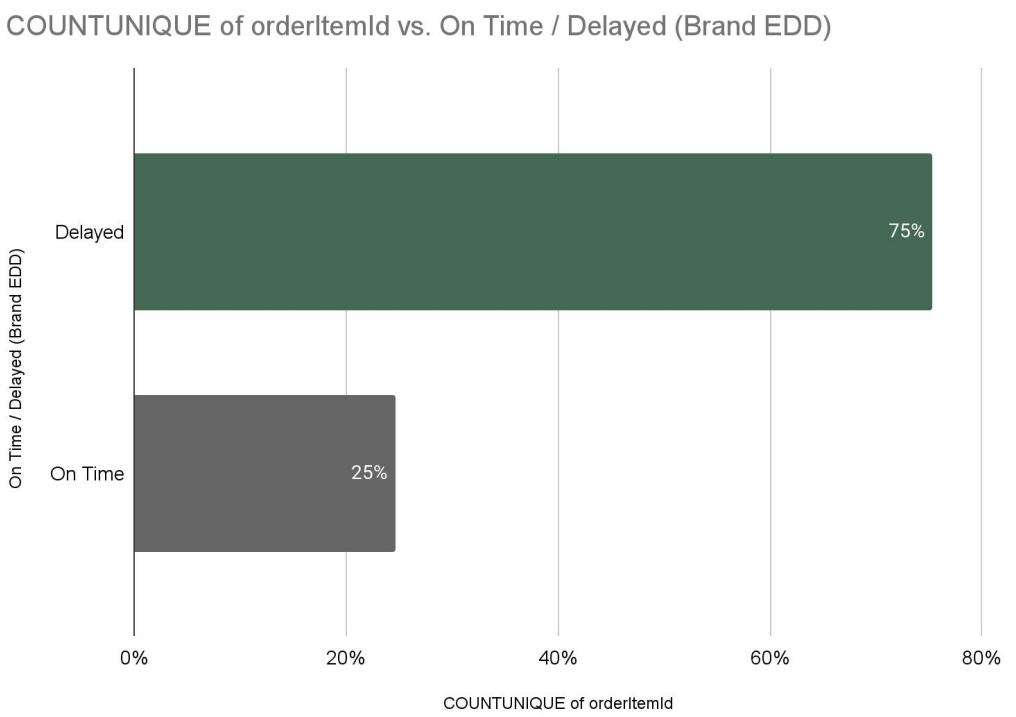
- 75% of Brand EDDs are inaccurate; compared to the average and could use a strong improvement
- 90% of our Top 10 customer's orders have missed both EDDs
- Funnel metrics remain stable; slight increase seen in quote requests in November
- Brand performance quality metrics can be designed & tracked (metrics can be designed leading to creating a brand performance scorecard dashboard) -> engage brands who are low performing / missing delivery estimates
- Brand loyalty & brand delivery performance can be catalysts to guiding repurchase decisions

Assumptions made for the purpose of the analysis

- Data is cleansed to remove orders / order_items with NULL dates
- Analysis focuses on deliveredOnTime != NULL
- More rows were cleansed where product Names = Test etc
- Summary:
  - 8515 unique order_item_ids were considered for the purpose of the analysis (from the order_item table & case_study table)
  - For delivery accuracy calculation:
    - It was assumed that when delivered_on_time > estimateddelivery startdate & delivered_on_time > estimateddeliveryenddate (ie delivery was missed within EDD window the order would be considered 'delayed')
  - Note: a category field was not found in the datasets (perhaps it was called something else / I've missed it?) - can speak to how to apply it here similar to brand / customer analyses
  - Risk: Data quality concerns can impact decisions / insights presented here

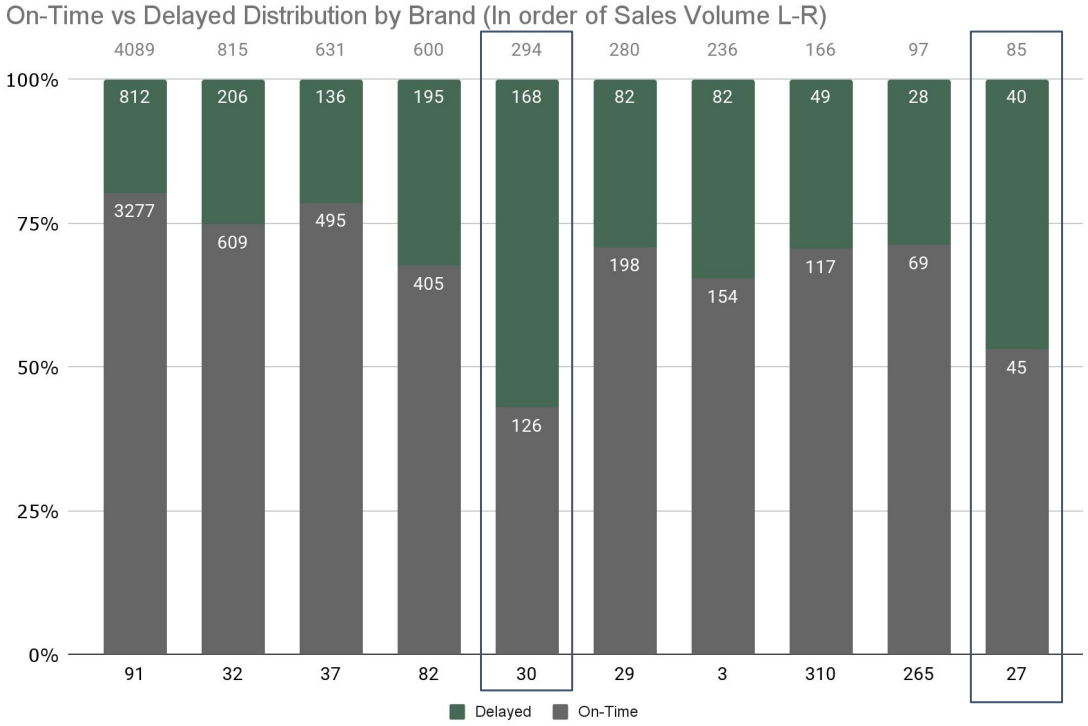
Q4: 75% of Brand EDDs are found to be inaccurate; with orders from brand 30 & 85 particularly

Brand EDD On Time vs. Delayed (RD



Note: Order is considered delayed if deliveredOnTime > EDDStart & deliveredOnTime > EDDEnd

On Time vs. Delayed Distribution by Brand (Top 10 brand)



- **While this could be a seasonal factor, this could have an effect on customers choosing to prefer alternative brands given their better delivery performance**

Q5: With most delays within the 0-15 day range; there's a lot of opportunity to align with brand partners to optimize customer experience

Near-term

- **Brand-specific analysis:**
  - Start with working with top 10-20 brands (by order size / revenue) to align on ways to improve delivery estimates. This could include building tooling, working with them to give them a larger lead time or exploring other fulfillment partners.
  - Conduct an assessment into historical accuracy, develop a scoring mechanism to provide "X expected" forecasts
- **Brand performance dashboard & strong feedback loop**: Design, define and track key brand quality & performance metrics into a brand performance dashboard. Brands with consistently delayed deliveries can cause customers to move to alternate platforms. Design / redesign processes to engage struggling brands. Share data with brand partners to enable them to improve their operational processes
- **Category analysis**: Do some categories perform worse than others? Conduct a deep-dive into order data across markets to determine best / worst performing categories
- **Alternatives**: Engage customers who've faced consistently late deliveries to explore other brand partners / alternatives

Longer term

API integration: Work with suppliers to develop / leverage APIs, connections to their OMS and building tooling / integrations to extract real-time info on order performance; will allow for near real-time information on order updates

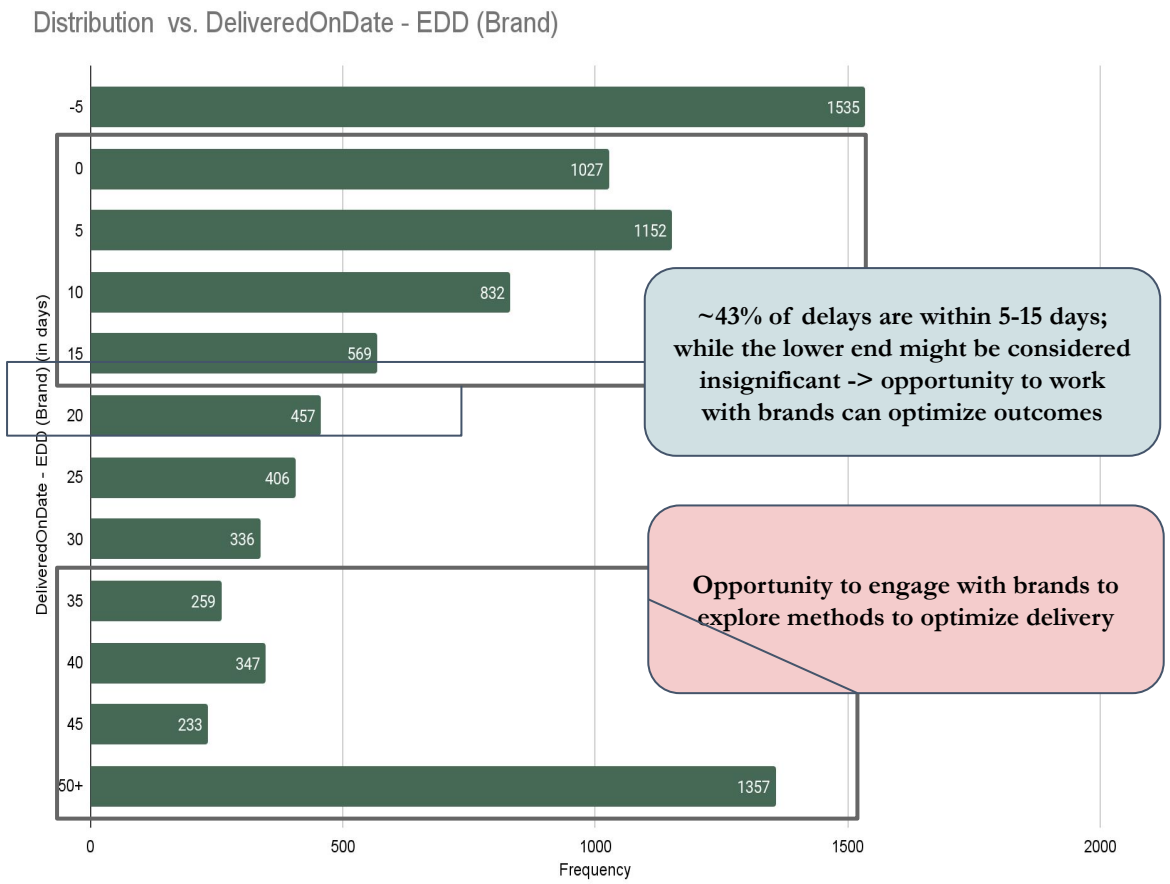
Order optimization based on warehouse location: Working with suppliers to optimize delivery based on closest warehouse

Intelligent forecasting:

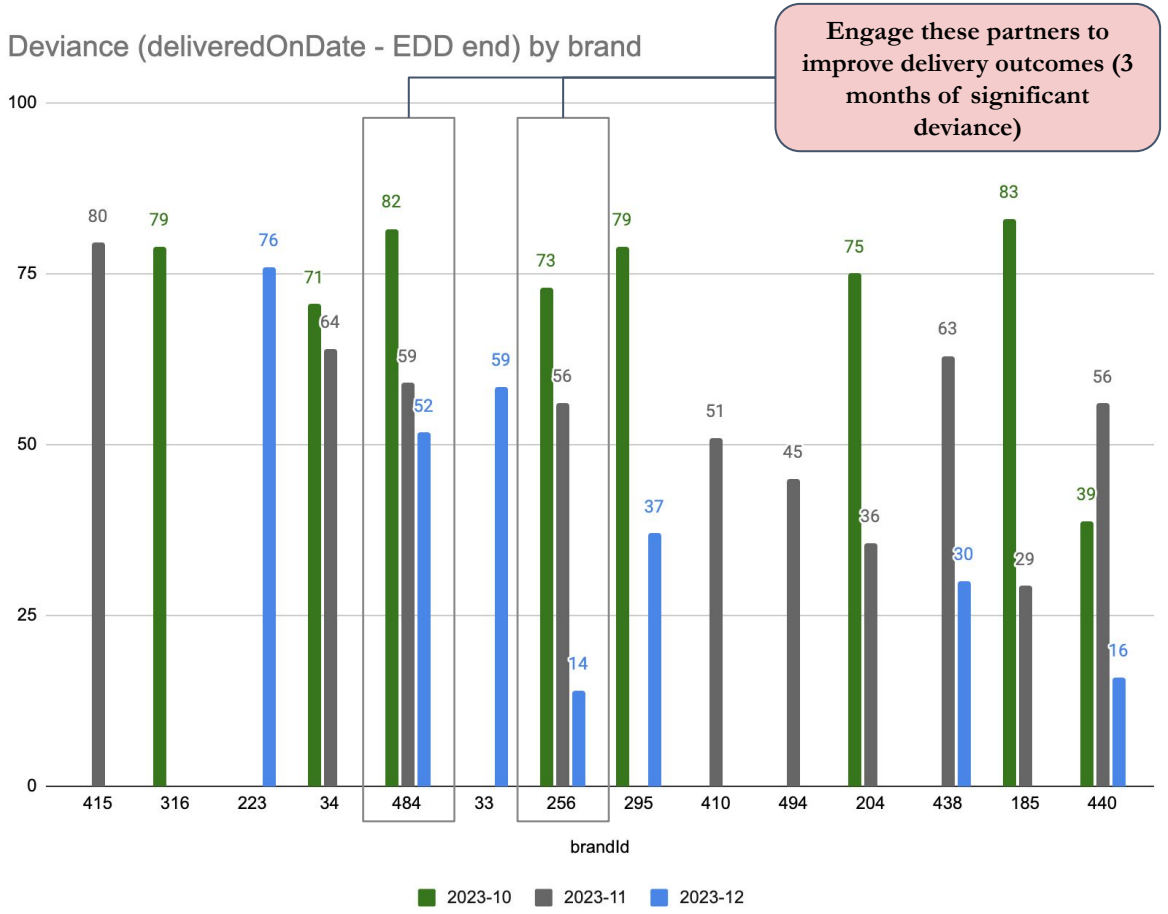
- Service categories which are unique may stand out and may experience higher demand and consequently drive revenue. We could use advanced analytics / ML methods & tools to forecast better customer demand & brand availability better - **use our data as a key asset to drive decision-making for brands**
- Post purchase / visit customer surveys can be leveraged; insight could be funneled to brands to help with inventory decisions (surveys could be incentivized to encourage participation)

Q5: With most delays within the 0-15 day range; there's a lot of opportunity to align with brand partners to optimize customer experience

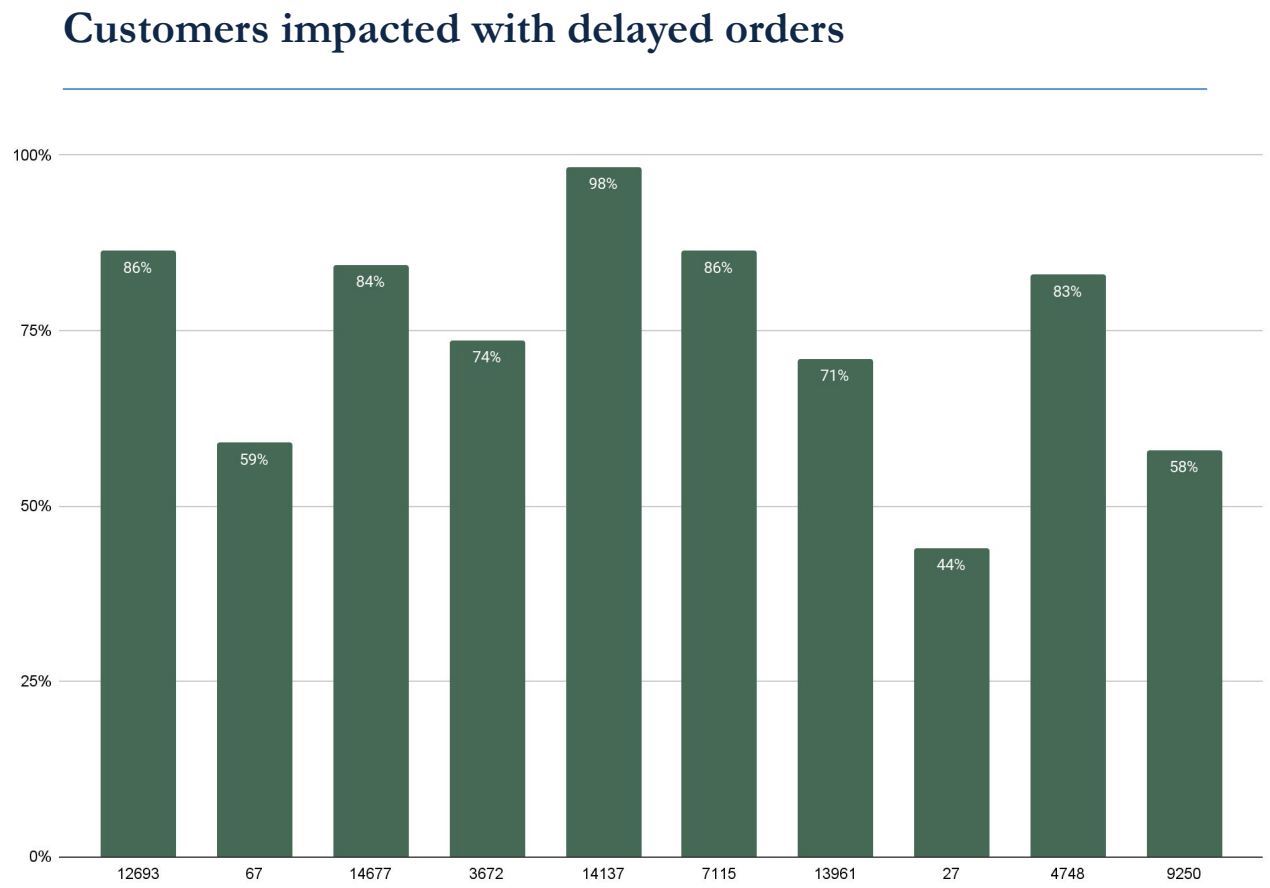
Frequency of occurrences by # of days past EDD



Avg. Deviance by brand



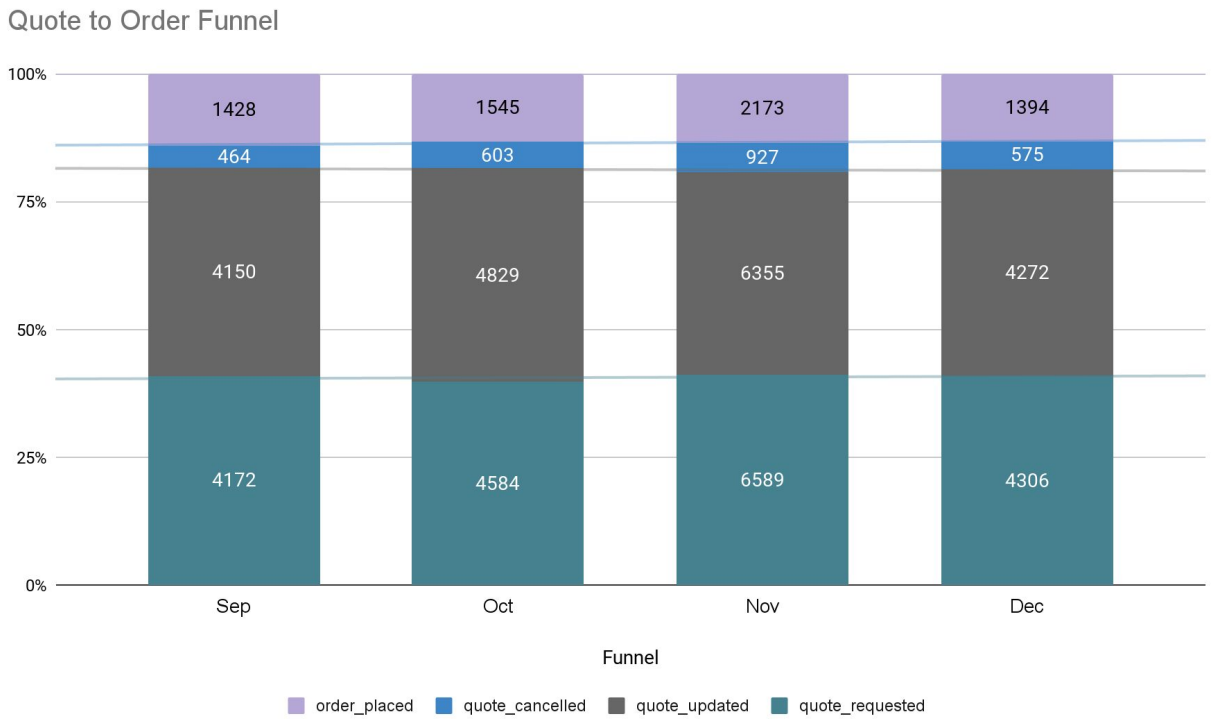
Q5: On the customer side however, we see 90% of our Top 10 customer's orders have missed both EDDs



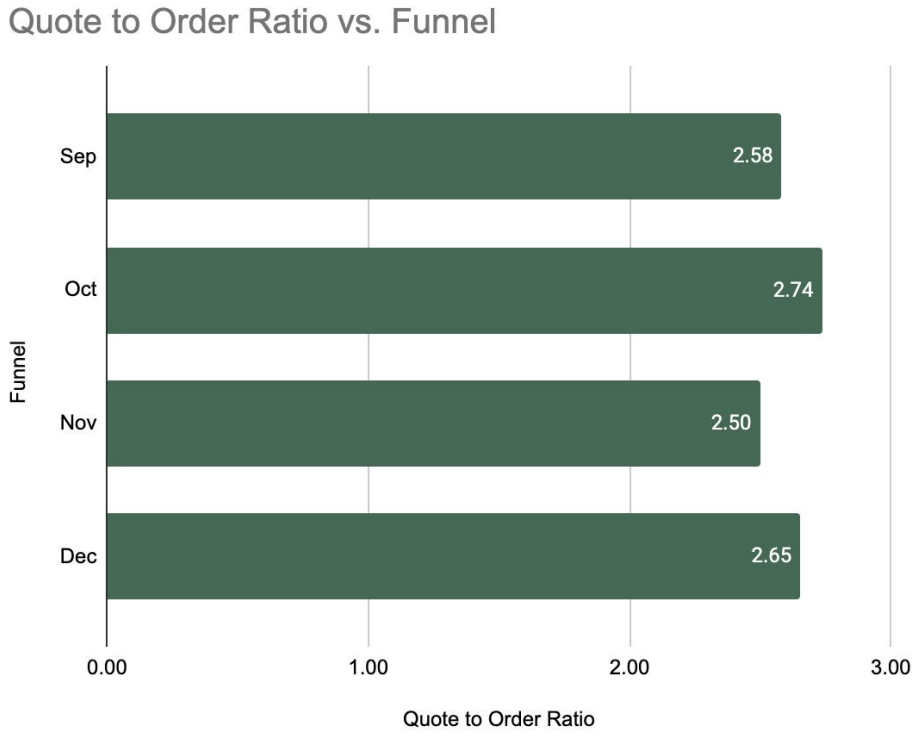
- **Particularly, customer 14137 & 12693 who seem to have ~90% of their orders delayed**
  - **This could have second order effects causing them to turn away from the platform or to other channels**

Q6: Funnel metrics remain stable; slight increase seen in quote requests in November

Brand EDD On Time vs. Delayed (when brands overshot)



On Time vs Delayed Distribution by Brand



- We see a relatively stable funnel metrics from Quote to Order (Quote_requested > Quote_updated > Quote_cancelled > Order placed)
- In November: While a slight increase (~2%) seen in quotes_requested (possibly due to holiday season); an 8% decrease in order conversions is observed (most likely due to cancellations possibly due to not being able to receive their items in time (send survey to investigate)

Q6: Visualization options for different periods: Line, Bar, Heatmaps

Daily / Weekly

Daily

- Line Chart: Helps visualize the daily fluctuations in order volume. This can help identify peak hours, anomalies or periods of low activity.
- Stacked Area Chart: Shows the distribution of different event types (e.g., order placed / processed) to summarize operating effectiveness

Weekly

- Bar Chart: Excellent for observing weekly patterns or trends. This can reveal if certain days of the week are busier or quieter for order processing
- Stacked Bar Chart: Show the distribution of different event types (e.g., order placed, order processed, order delivered) for each month to understand the order processing workflow and any changes in the distribution over time

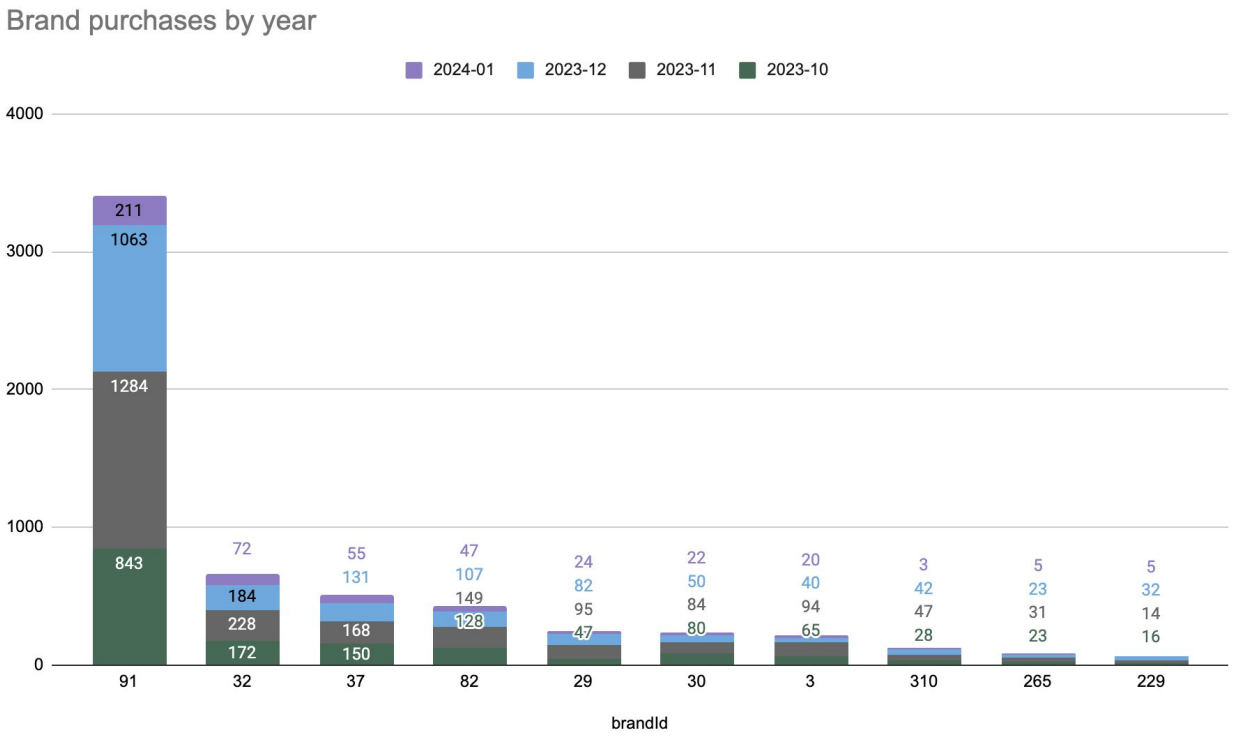
Monthly

Monthly

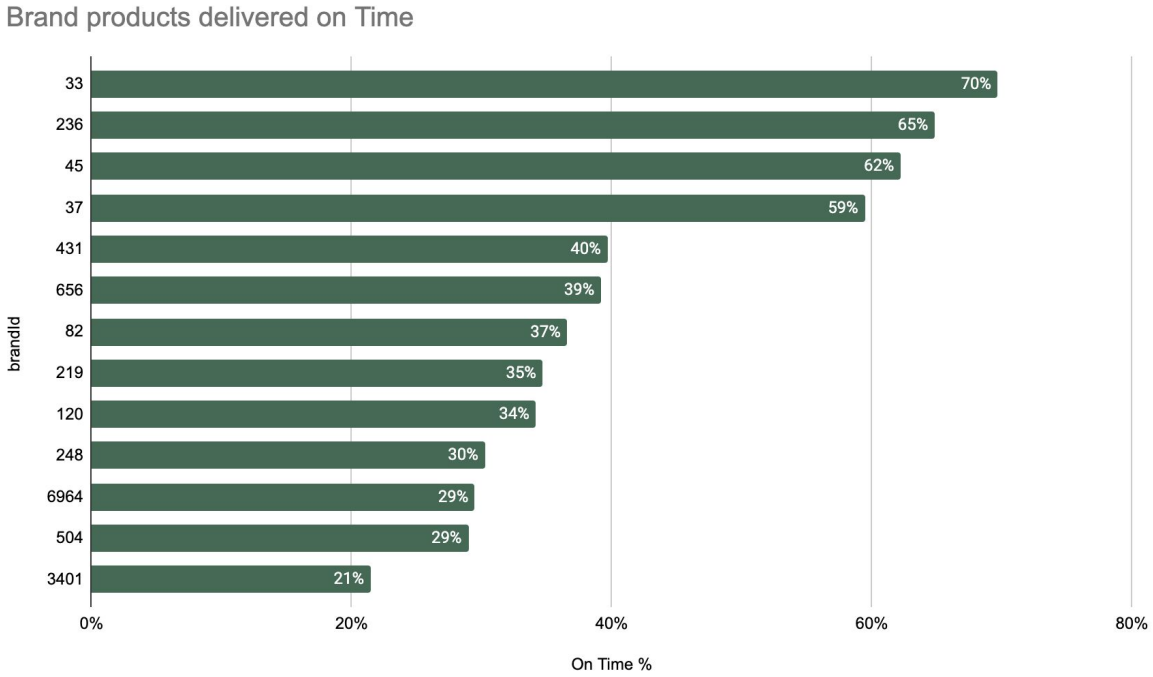
- Line Chart or Bar Chart: Plot the total number of orders over each month to visualize monthly trends in order volume. This can help identify seasonal variations or growth trends over time.
- Stacked Bar Chart: Show the distribution of different event types (e.g., order placed, order processed, order delivered) for each month to understand the order processing workflow and any changes in the distribution over time.
- Heatmap on a map: Heatmaps overlaid on maps are ideal for visualizing monthly data, offering intuitive insights into spatial trends and variations over time. Their color intensity allows for quick identification of patterns, while interactive features enable easy exploration and comparison between different months.
-

Q7: Two possible hypotheses purchases by future customers

Most bought brands will be repurchased by customers



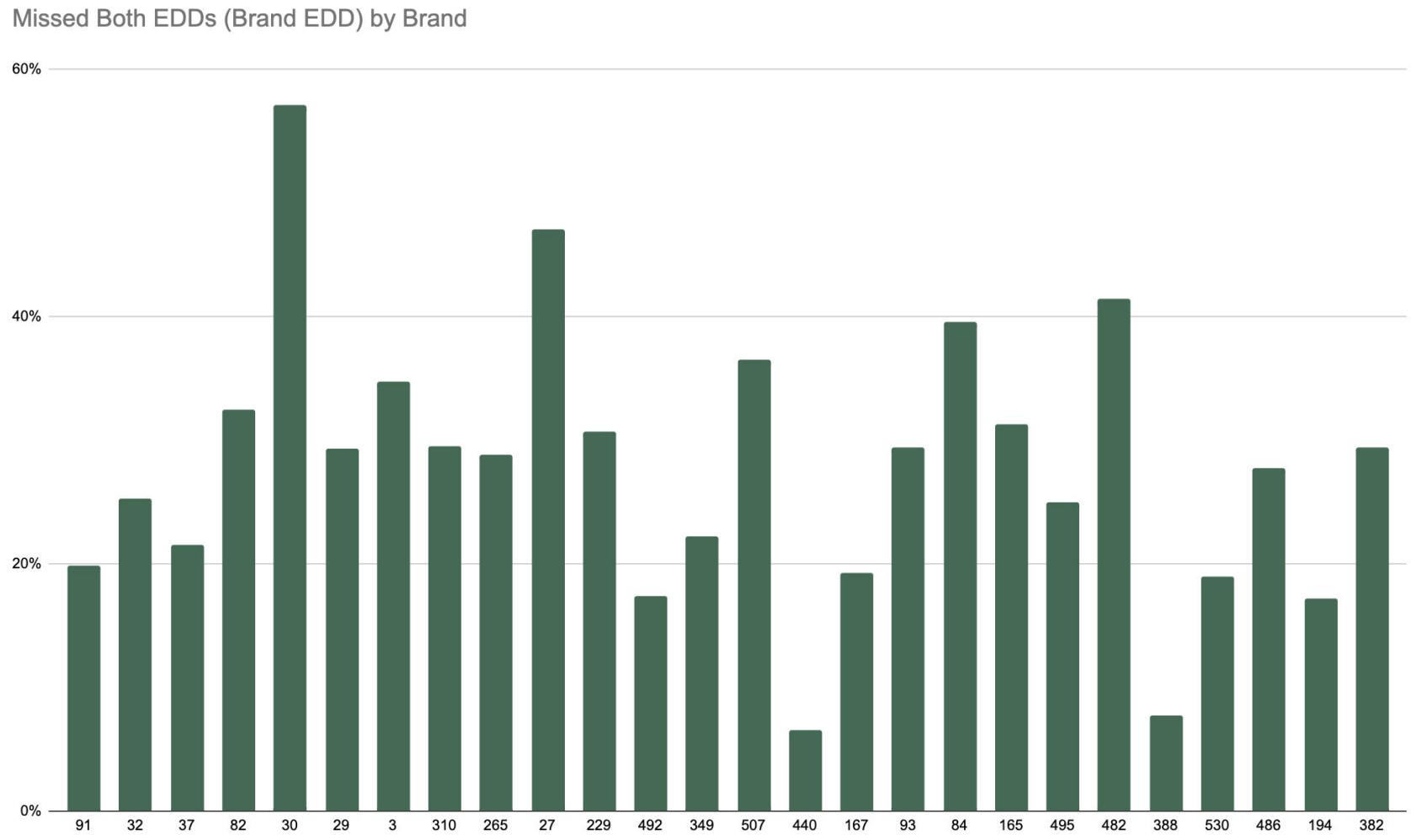
Brand products delivered on Time



- Brands 91, 32 & 37 seem to be preferred by buyers given their high purchase frequency (in terms of items purchased and spend)
- However, customers may prefer 37 over the others if on time delivery is a factor -> 37 tends to have a much higher % of its products delivered within its specified delivery window (Cancellation % may be another useful metric)

Appendix

Appendix A: Distribution of brands who missing both EDDs



Appendix B: Use predictive modeling to predict delivery accuracy

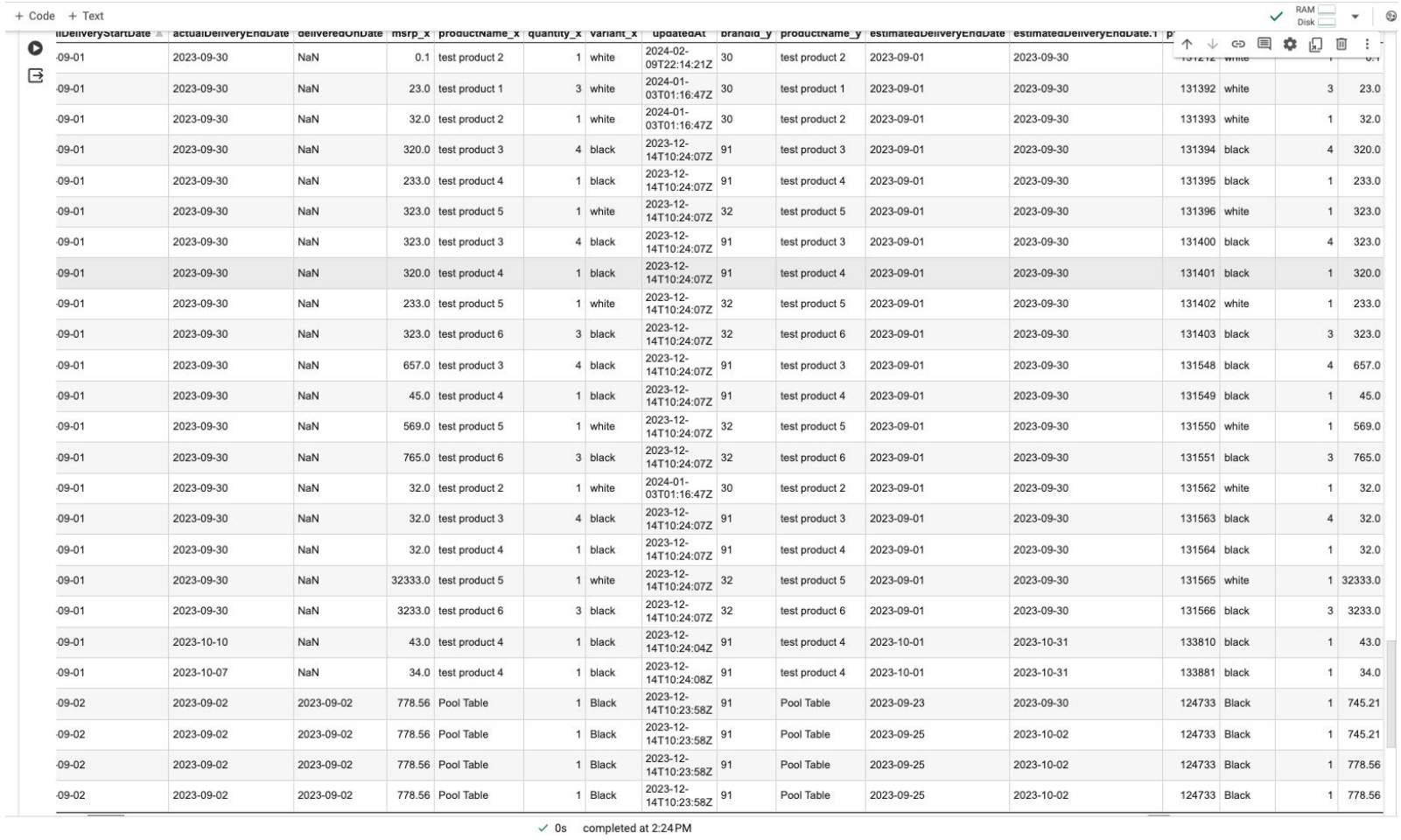
```
[ ]
import pandas as pd
import statsmodels.api as sm


# Drop rows with NaN values in date columns
df.dropna(subset=['deliveredOnDate', 'actualDeliveryEndDate', 'actualDeliveryStartDate', 'estimatedDeliveryStartDate', 'estimatedDeliveryEndDate'], inplace=True)

# Select features for prediction
features = ['totalCost', 'deliveryFees', 'deltaStartDate', 'deltaEndDate', 'deliveredOnDate-actualDeliveryEndDate']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(df[features], df['AccurateDelivery'], test_size=0.25, random_state=42)

# Train the Logistic Regression model
model = sm.Logit(y_train, X_train)
result = model.fit()

# Get the p-values of the model coefficients
p_values = result.pvalues

# Select features with p-value less than 0.05
significant_features = p_values[p_values < 0.05].index.tolist()

# Print the significant features
print("Significant features:", significant_features)

# Print the p-value of the model
print("Model p-value:", result.pvalues[0])

# Evaluate the model's accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
```

```
Optimization terminated successfully.
         Current function value: 0.619807
         Iterations 6
Significant features: ['totalCost', 'deliveryFees', 'deltaStartDate', 'deltaEndDate', 'deliveredOnDate-actualDeliveryEndDate']
Model p-value: 3.7647047523604935e-08
Accuracy: 0.7524659464537341
```

Interpretation

- The log reg results show that:
  - totalCost, fees and estimated dates are a good predictors for accurate Delivery
  - the model here has an accuracy of 75%
- The **p-values for features are less than 0.05,** which indicates that there is a **statistically significant** relationship between fees, cost & date.

Appendix C: Data quality concerns:



+ Code + Text

lDeliveryStartDate	actualDeliveryEndDate	deliveredOnDate	msrp_x	productName_x	quantity_x	variant_x	updatedAt	brandId_y	productName_y	estimatedDeliveryEndDate	estimatedDeliveryEndDate.1	p			
-09-01	2023-09-30	NaN	0.1	test product 2	1	white	2024-02-09T22:14:21Z	30	test product 2	2023-09-01	2023-09-30	131212	white	1	0.1
-09-01	2023-09-30	NaN	23.0	test product 1	3	white	2024-01-03T01:16:47Z	30	test product 1	2023-09-01	2023-09-30	131392	white	3	23.0
-09-01	2023-09-30	NaN	32.0	test product 2	1	white	2024-01-03T01:16:47Z	30	test product 2	2023-09-01	2023-09-30	131393	white	1	32.0
-09-01	2023-09-30	NaN	320.0	test product 3	4	black	2023-12-14T10:24:07Z	91	test product 3	2023-09-01	2023-09-30	131394	black	4	320.0
-09-01	2023-09-30	NaN	233.0	test product 4	1	black	2023-12-14T10:24:07Z	91	test product 4	2023-09-01	2023-09-30	131395	black	1	233.0
-09-01	2023-09-30	NaN	323.0	test product 5	1	white	2023-12-14T10:24:07Z	32	test product 5	2023-09-01	2023-09-30	131396	white	1	323.0
-09-01	2023-09-30	NaN	323.0	test product 3	4	black	2023-12-14T10:24:07Z	91	test product 3	2023-09-01	2023-09-30	131400	black	4	323.0
-09-01	2023-09-30	NaN	320.0	test product 4	1	black	2023-12-14T10:24:07Z	91	test product 4	2023-09-01	2023-09-30	131401	black	1	320.0
-09-01	2023-09-30	NaN	233.0	test product 5	1	white	2023-12-14T10:24:07Z	32	test product 5	2023-09-01	2023-09-30	131402	white	1	233.0
-09-01	2023-09-30	NaN	323.0	test product 6	3	black	2023-12-14T10:24:07Z	32	test product 6	2023-09-01	2023-09-30	131403	black	3	323.0
-09-01	2023-09-30	NaN	657.0	test product 3	4	black	2023-12-14T10:24:07Z	91	test product 3	2023-09-01	2023-09-30	131548	black	4	657.0
-09-01	2023-09-30	NaN	45.0	test product 4	1	black	2023-12-14T10:24:07Z	91	test product 4	2023-09-01	2023-09-30	131549	black	1	45.0
-09-01	2023-09-30	NaN	569.0	test product 5	1	white	2023-12-14T10:24:07Z	32	test product 5	2023-09-01	2023-09-30	131550	white	1	569.0
-09-01	2023-09-30	NaN	765.0	test product 6	3	black	2023-12-14T10:24:07Z	32	test product 6	2023-09-01	2023-09-30	131551	black	3	765.0
-09-01	2023-09-30	NaN	32.0	test product 2	1	white	2024-01-03T01:16:47Z	30	test product 2	2023-09-01	2023-09-30	131562	white	1	32.0
-09-01	2023-09-30	NaN	32.0	test product 3	4	black	2023-12-14T10:24:07Z	91	test product 3	2023-09-01	2023-09-30	131563	black	4	32.0
-09-01	2023-09-30	NaN	32.0	test product 4	1	black	2023-12-14T10:24:07Z	91	test product 4	2023-09-01	2023-09-30	131564	black	1	32.0
-09-01	2023-09-30	NaN	32333.0	test product 5	1	white	2023-12-14T10:24:07Z	32	test product 5	2023-09-01	2023-09-30	131565	white	1	32333.0
-09-01	2023-09-30	NaN	3233.0	test product 6	3	black	2023-12-14T10:24:07Z	32	test product 6	2023-09-01	2023-09-30	131566	black	3	3233.0
-09-01	2023-10-10	NaN	43.0	test product 4	1	black	2023-12-14T10:24:04Z	91	test product 4	2023-10-01	2023-10-31	133810	black	1	43.0
-09-01	2023-10-07	NaN	34.0	test product 4	1	black	2023-12-14T10:24:08Z	91	test product 4	2023-10-01	2023-10-31	133881	black	1	34.0
-09-02	2023-09-02	2023-09-02	778.56	Pool Table	1	Black	2023-12-14T10:23:58Z	91	Pool Table	2023-09-23	2023-09-30	124733	Black	1	745.21
-09-02	2023-09-02	2023-09-02	778.56	Pool Table	1	Black	2023-12-14T10:23:58Z	91	Pool Table	2023-09-25	2023-10-02	124733	Black	1	745.21
-09-02	2023-09-02	2023-09-02	778.56	Pool Table	1	Black	2023-12-14T10:23:58Z	91	Pool Table	2023-09-25	2023-10-02	124733	Black	1	778.56
-09-02	2023-09-02	2023-09-02	778.56	Pool Table	1	Black	2023-12-14T10:23:58Z	91	Pool Table	2023-09-25	2023-10-02	124733	Black	1	778.56

✓ 0s completed at 2:24PM