

Analytics Practicum

Interim Presentation



Agenda

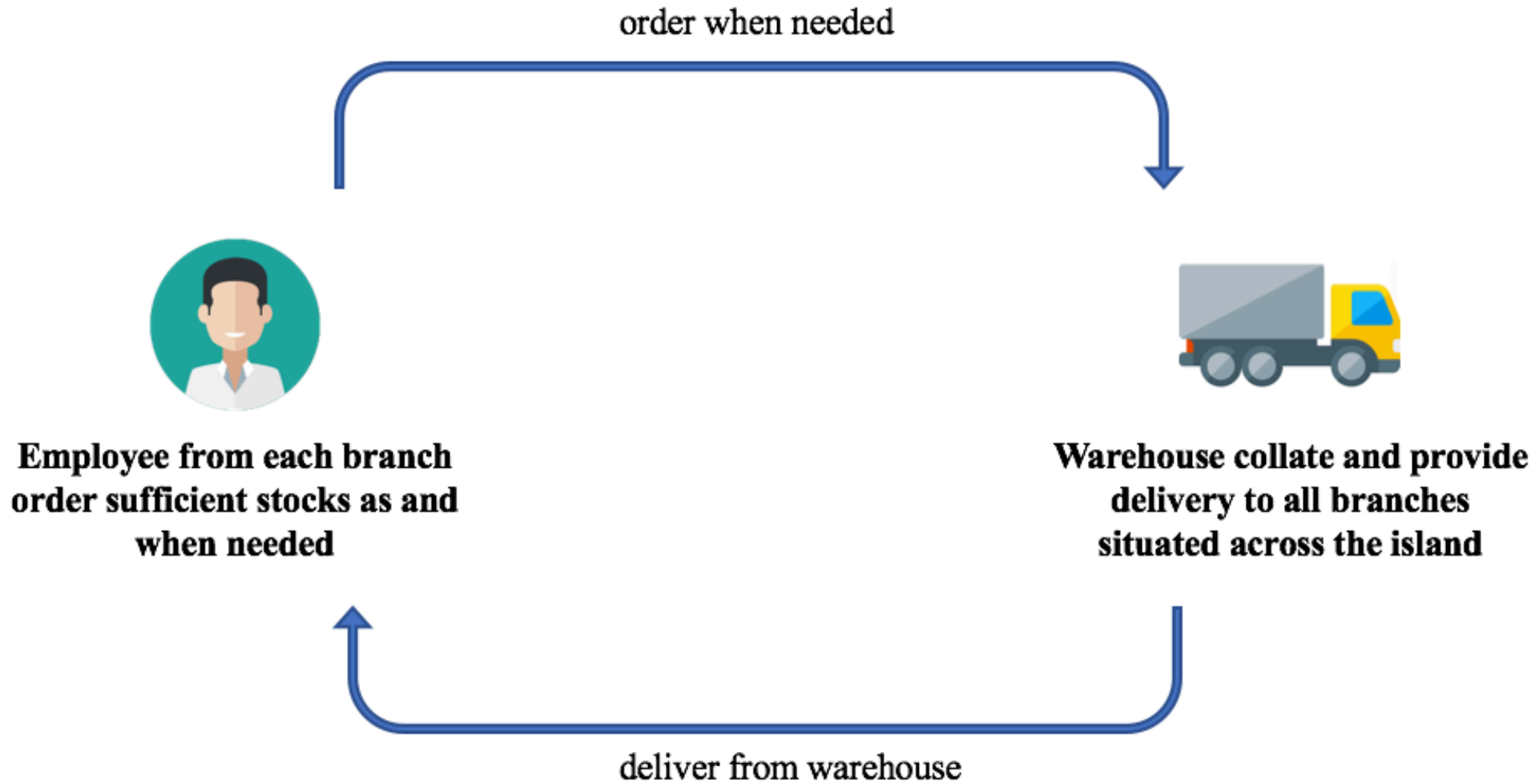
- Overview
- Data Integration and filtering
- Data Cleaning
- Data Preparation
- Exploratory Data Analysis Findings
- Revised Scope of Work
- Revised Work Plan
- Bibliography

Sponsor Background

- Most well-known bubble tea brands in Singapore
- Origin from Taiwan
- Officially launched in Singapore in 2007
- Total of 46 outlets widespread in Singapore



Current Re-order Model



Project Objectives

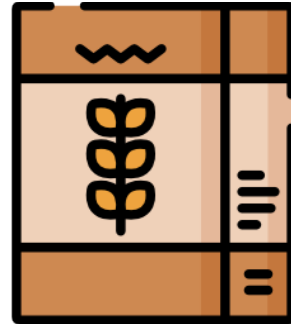
Reorder Optimal Stock Quantity

- To allocate the optimal restock amount
 - *How much should employee order to ensure that they will not run out of ingredients needed?*
- To analyze the time series trend
 - *How much should the employee order in this current period?*

Data Collection



47 Outlets



224 Products



2 Years Data
[Jan 2016 - Dec 2017]

Data Collection



Delivery
Data



Promotion
Data



Outlets
Data

Data Cleaning

- Redundant columns – Delivery Data

Column name	Description
Delivery Date	The date where the order is being delivered to the store.
Created Date	The data where the order is being placed by the staff.
Time	The time where the order is being placed by the staff
Order Num	The unique id of the order.
Type	The different type of order being placed.
Store	The store outlet name and code that placed the order.
Product Num	The unique id of the product being ordered.
Product	The name of the product being ordered.
UOM	The unit of measures of the delivery order type. For instance: bottles, box, bag and roll.
Qty	The amount of product being placed.

Data Cleaning

- Noise
 - Oversea Branches order
 1. KOI TE (THAILAND)
 2. Cambodia Karanak KOI Café
 3. Jakarta
 4. Koi Myanmar
 5. Vietnam – Cong TY TNHH KOI NHANH
 - Closed down outlets
 1. Esplanade (May 2017)
 - Remove 2015 and 2018 data

Data Preparation

New Columns Created	Reason for creating the column
Created Date	Reason for creating a new column created date was due to the date format that KOl provided.
Date Diff	Data difference illustrates what is the difference in number of days for this outlet current order against the previous order.
Actual Date Diff	Actual date difference helps to standardize the difference in number of days for this outlet current order against the previous order across for all products in a single order

Data Preparation

New Columns Created

Reason for creating the column

Product Category

Reason for creating a Product Category column was because there was no significant segregation for the items.

Ingredients refer to bubble tea ingredients, non-ingredients refer to non-bubble tea ingredients but are crucial to business operation and lastly miscellaneous for others. For illustration purpose, examples on how we segregate the products are as follow:

- **Ingredients** → Pearl, Milk Powder, Jasmine Tea etc.
- **Non-Ingredients** → Cups, Straw, Cup Seal, Plastic Bags etc.
- **Miscellaneous** → Coins, Cleaning Supplies, Uniform etc.

Data Preparation

New Columns Created

Reason for creating the column

Include?

Reason for creating the include column was for our future model building. As there are some products identified with non-significant time-series trends, include will allow us to segregate products that have time-series trends and products without.

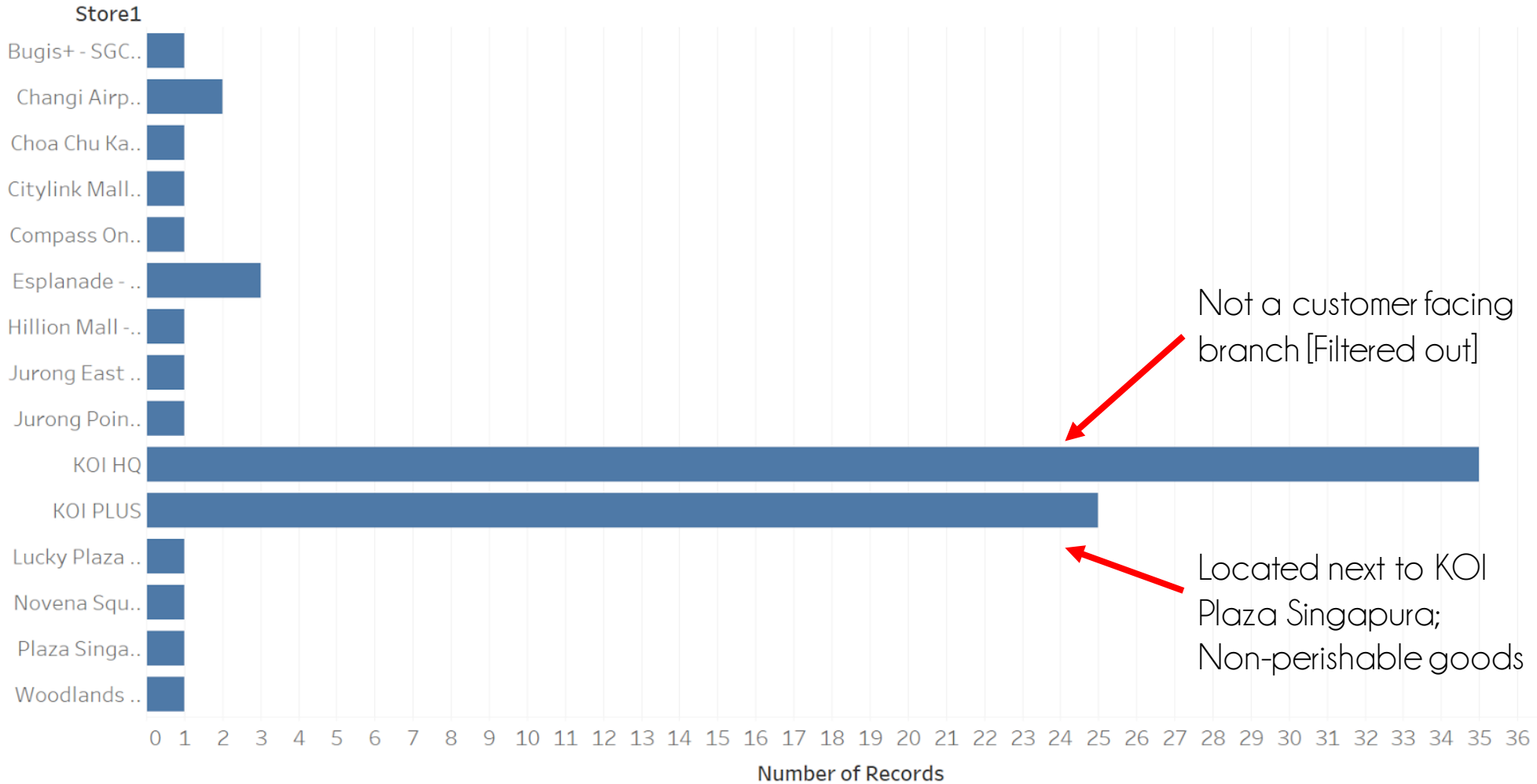
Those with time-series trend, it will be marked "Yes". Those without will be marked with "No"

Data Preparation

New Columns Created	Reason for creating the column
Segment	Reason for creating a new column "Segment" was because we wish to split the outlets with similar reorder frequency into groups to compare and make necessary business recommendations to our sponsor. We have split the outlets into 3 segments - Segment 1 for Reordering Frequency of ≤ 5 days, Segment 2 for Reordering Frequency of ≤ 9 days, and Segment 3 for Reordering Frequency with outliers of ≥ 10 days.
Region	Reason for creating a new column "Region" was to split our outlets into different constituency region in Singapore to enable us to provide better business recommendation regarding delivery and reorder frequency to our sponsor

Data Preparation

Store with ≥ 7 order date different



Data Preparation

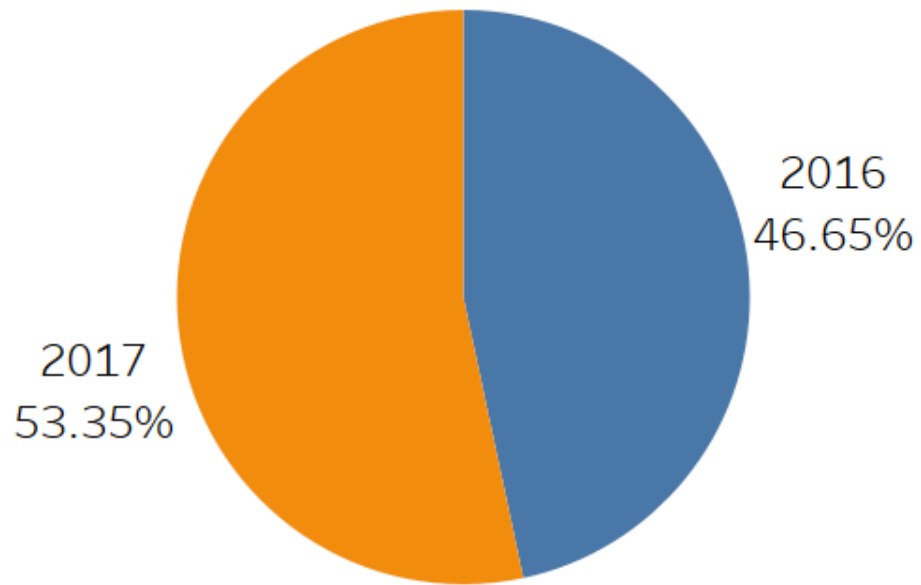
- Discontinued of ingredients
 1. Mango Syrup
 2. Mango Sticker
 3. Passion Fruit Syrup
 4. Passion Fruit Sticker
 5. Coffee Mate 1KG

Exploratory Data Analysis

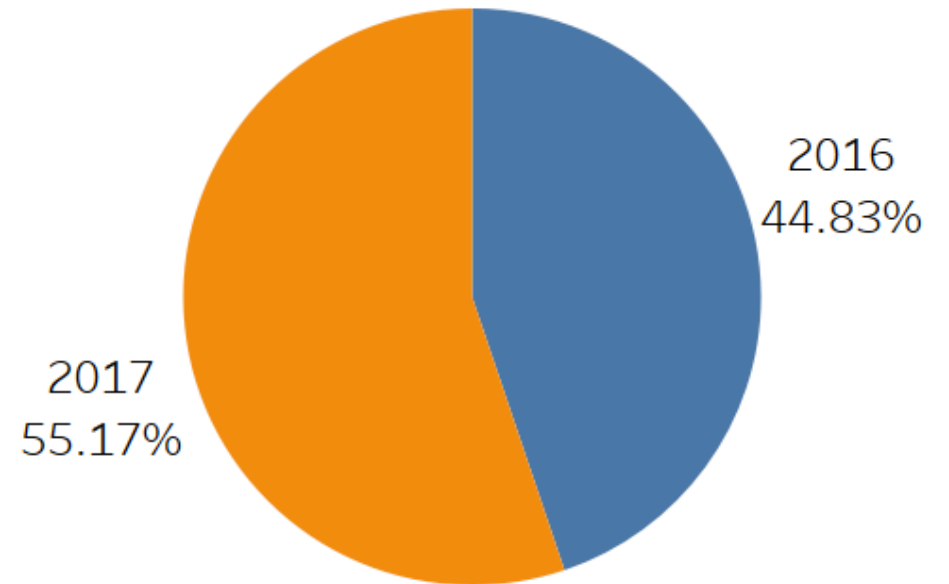
- 2016/2017 Delivery Data Analysis
- New Outlets Data Analysis
- Top 15 Outlets Data Analysis
- Product Category Outlet Analysis
- Top 15 Products Data Analysis
- Reordering Frequency Data Analysis
- Reordering Frequency Outliers Analysis

Exploratory Data Analysis

Percentage of total delivery by year

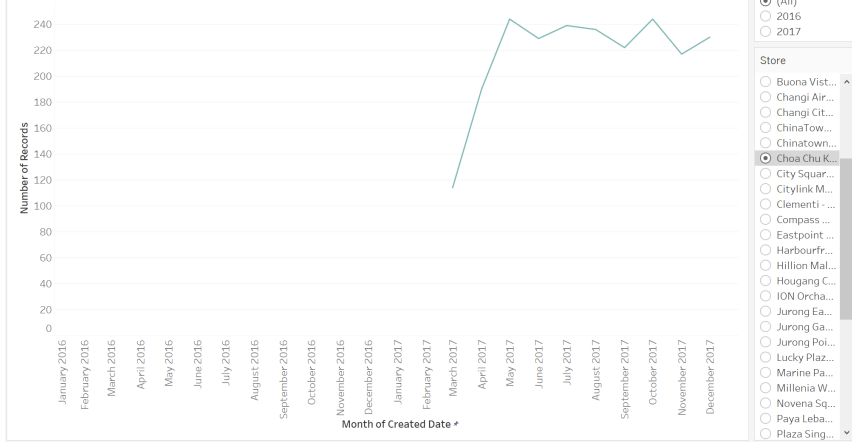


Percentage of total quantity by year

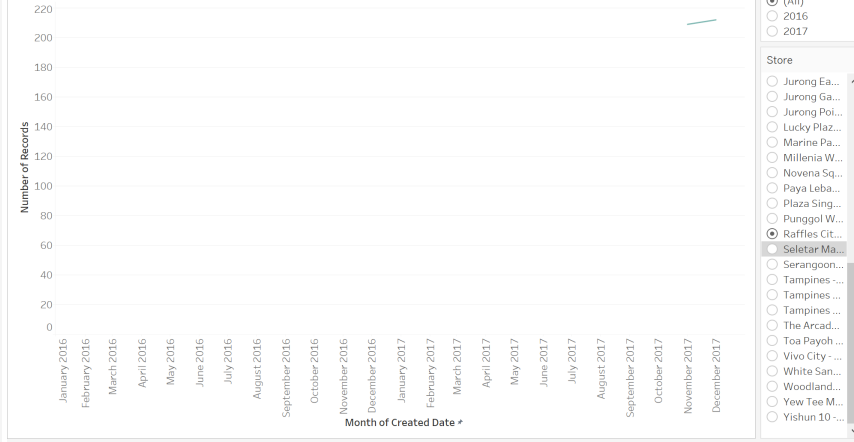


Exploratory Data Analysis

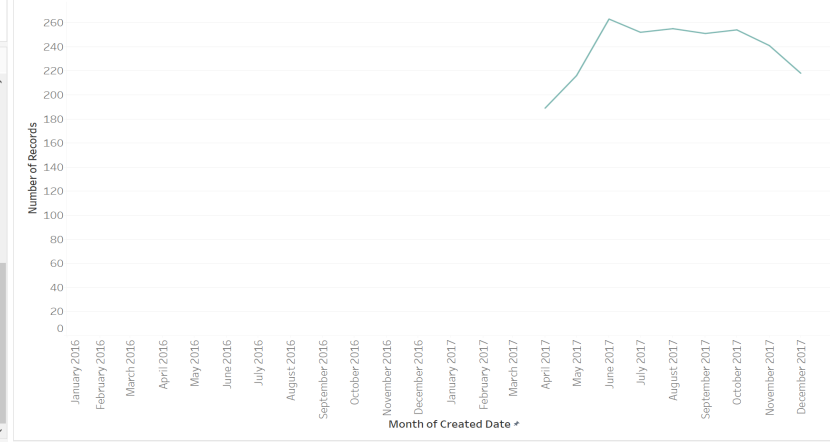
Number of records over the years for Choa Chu Kang MRT - SGE013



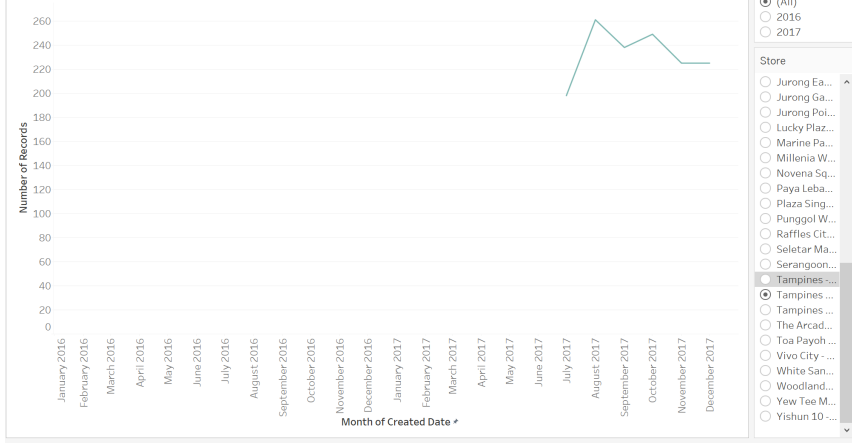
Number of records over the years for Raffles City - SGE014



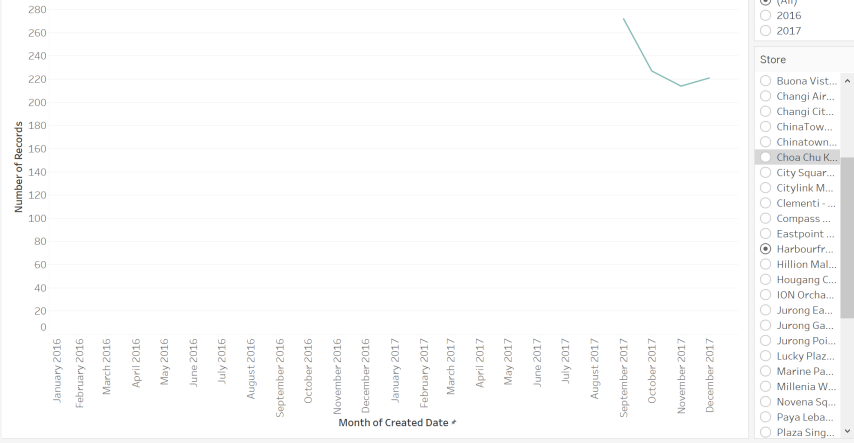
Number of records over the years for Novena Square - SGC031



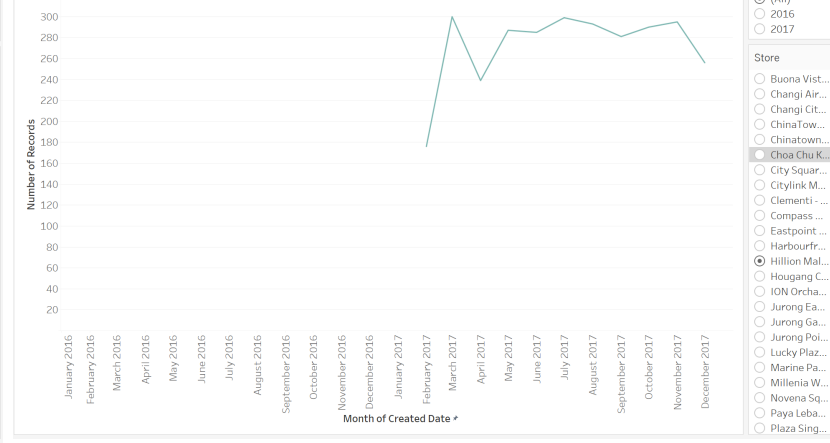
Number of records over the years for Tampines Mall - SGC032



Number of records over the years for Harbourfront - SGC033



Number of records over the years for Hillion Mall - SGC030



Factors affect the increment

- New outlets opening in 2016 & 2017

Opening in 2016

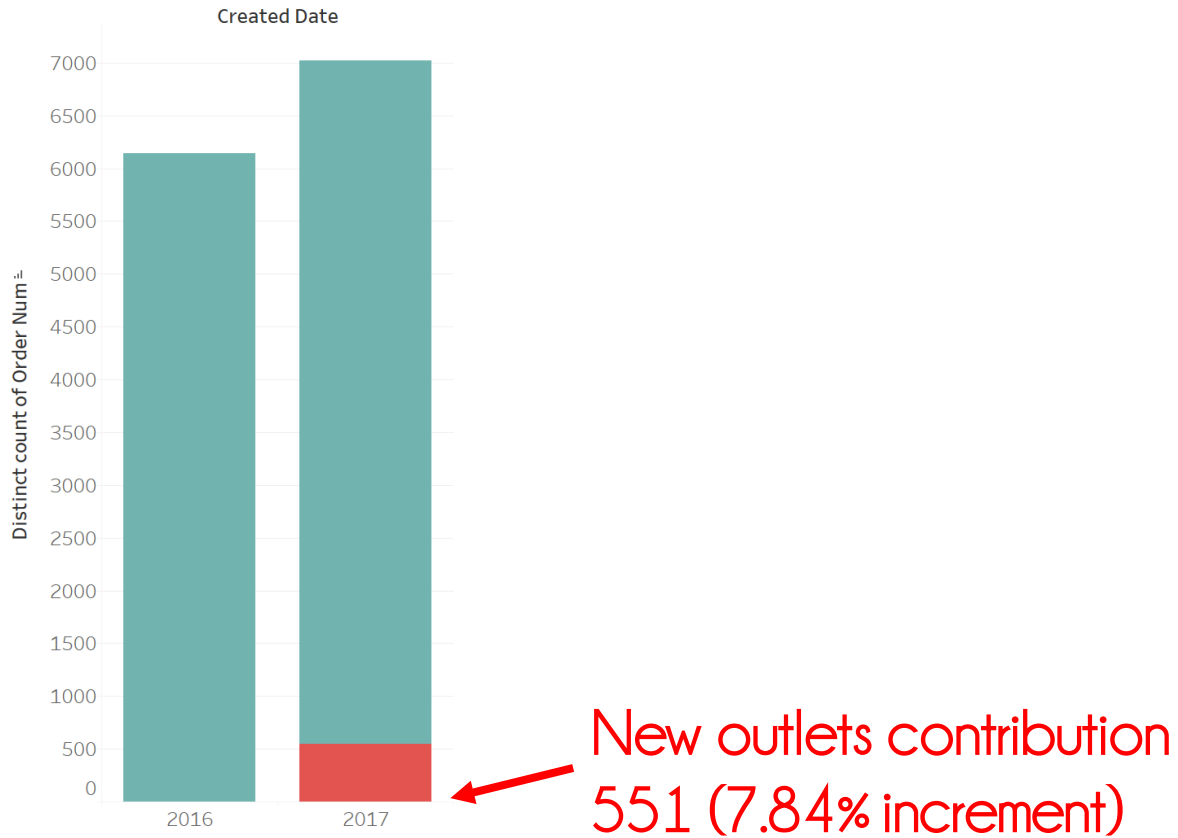
- Woodlands MRT (May 2016)
- CompassPoint (September 2016)
- The Arcade (November 2016)
- Jurong Point (December 2016)

Opening in 2017

- Harbourfront (Sept 2017)
- Jurong East (Nov 2017)
- Raffles City (Nov 2017)
- Hillion Mall (Feb 2017)
- CCK (April 2017)
- Novena Square (April 2017)
- Tampines Mall (July 2017)

Exploratory Data Analysis

Contribution of new outlets by delivery

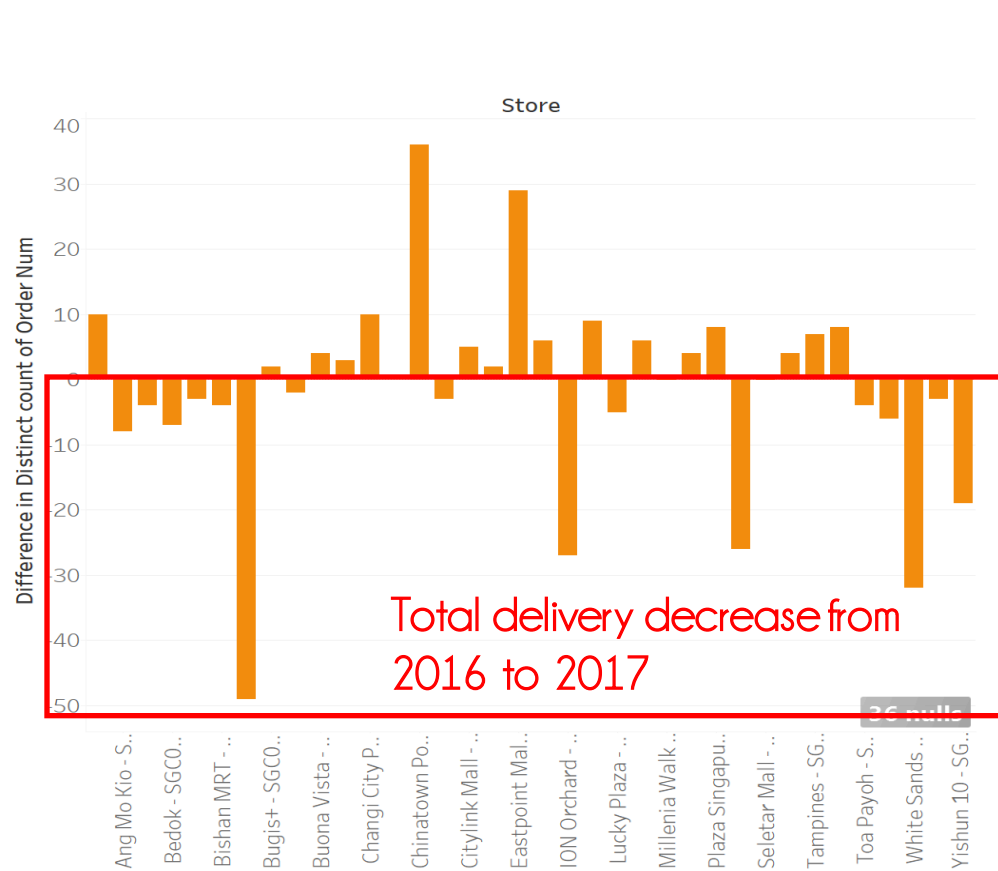


Contribution of new outlets by quantity

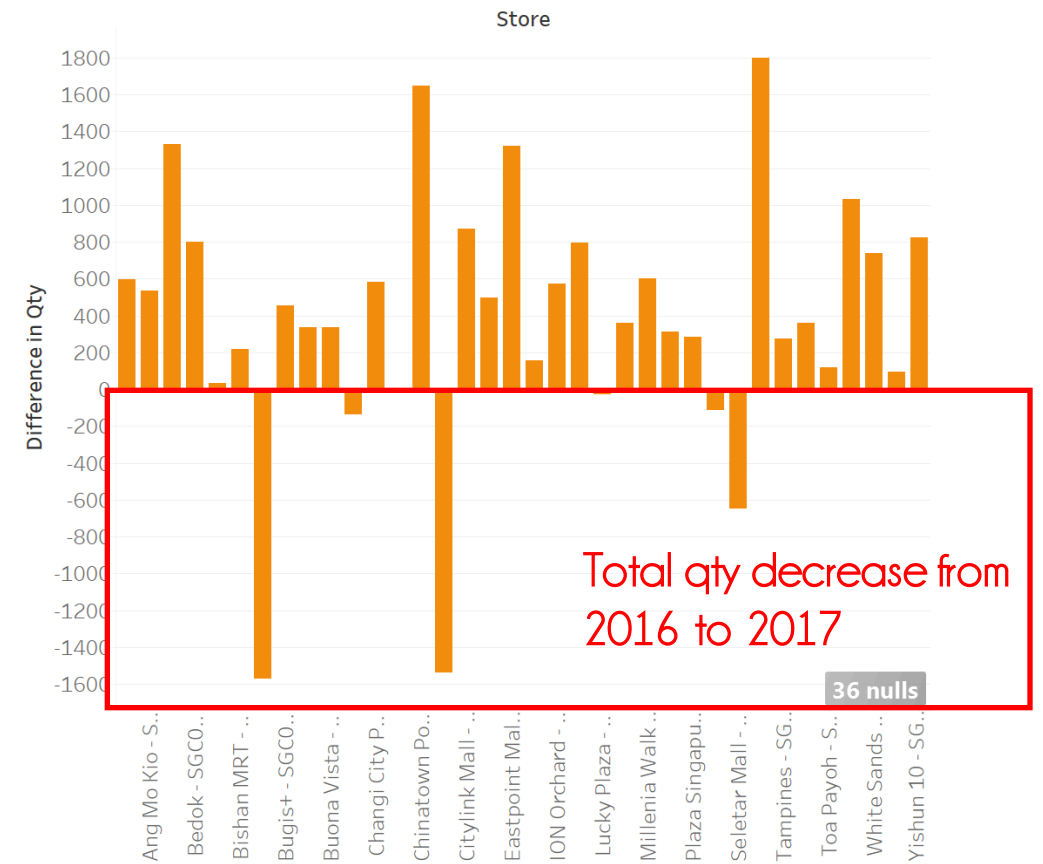


Exploratory Data Analysis

Contribution of existing outlets by delivery

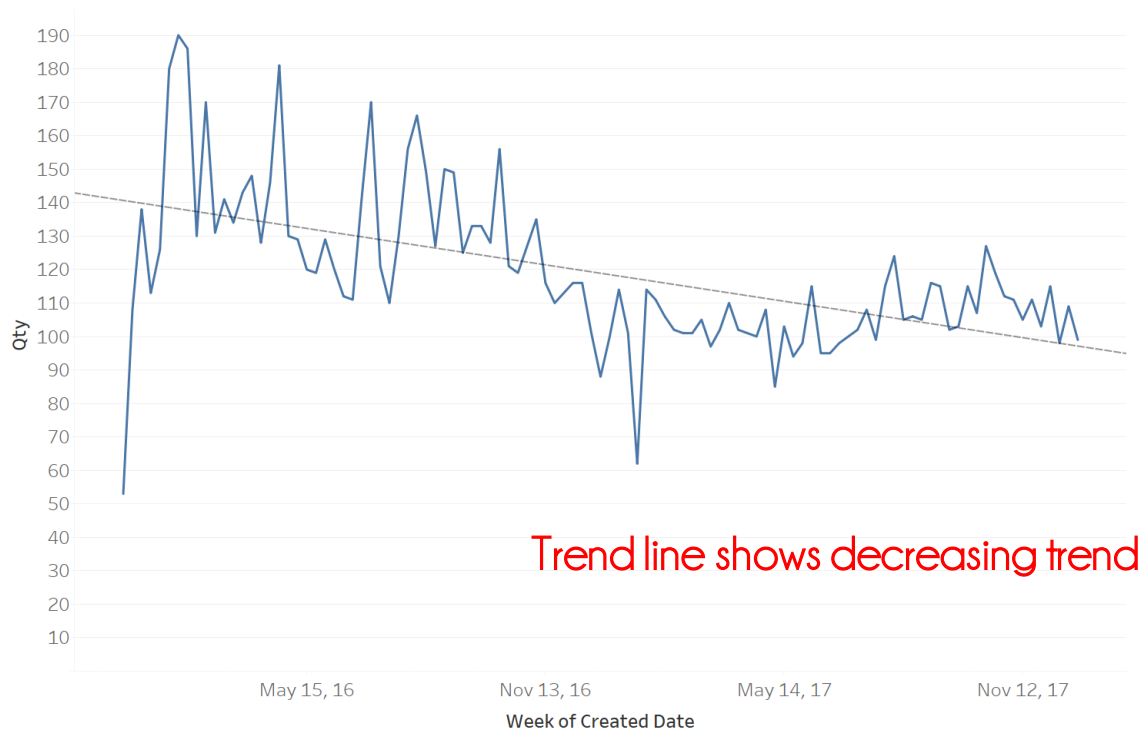


Contribution of existing outlets by quantity

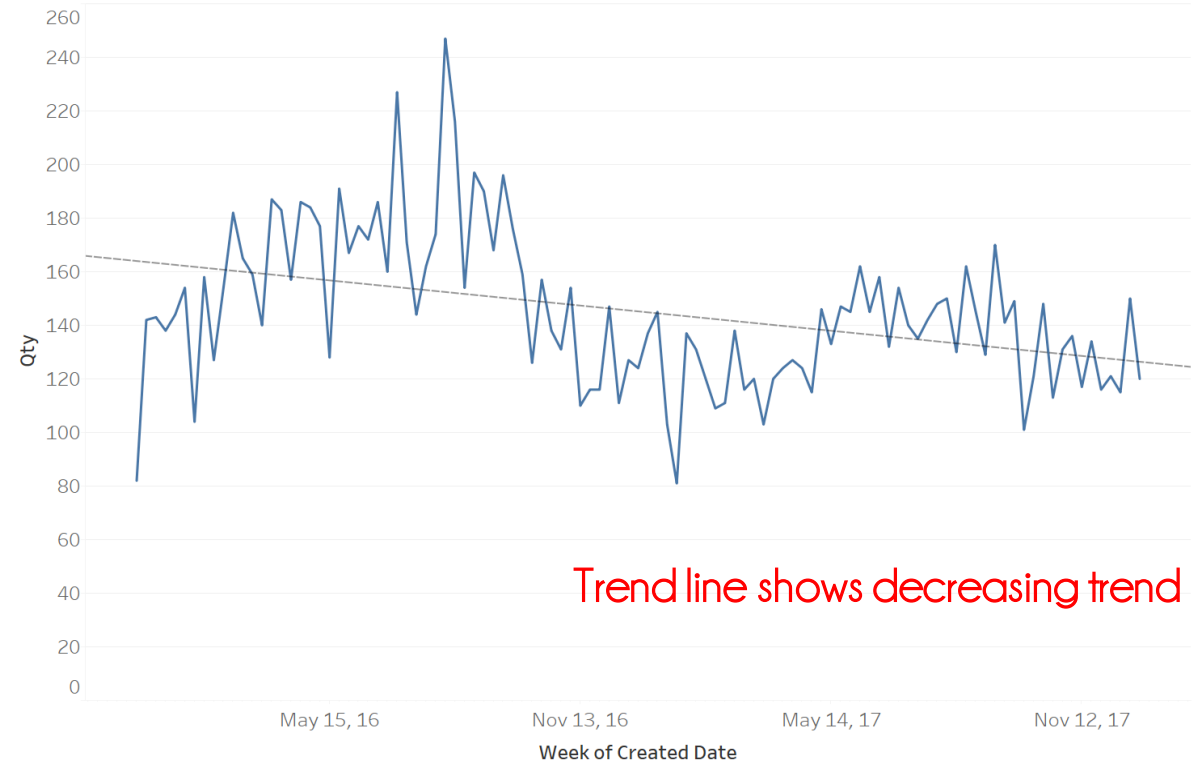


Exploratory Data Analysis

Time series analysis on the quantity ordered for Boon Lay MRT

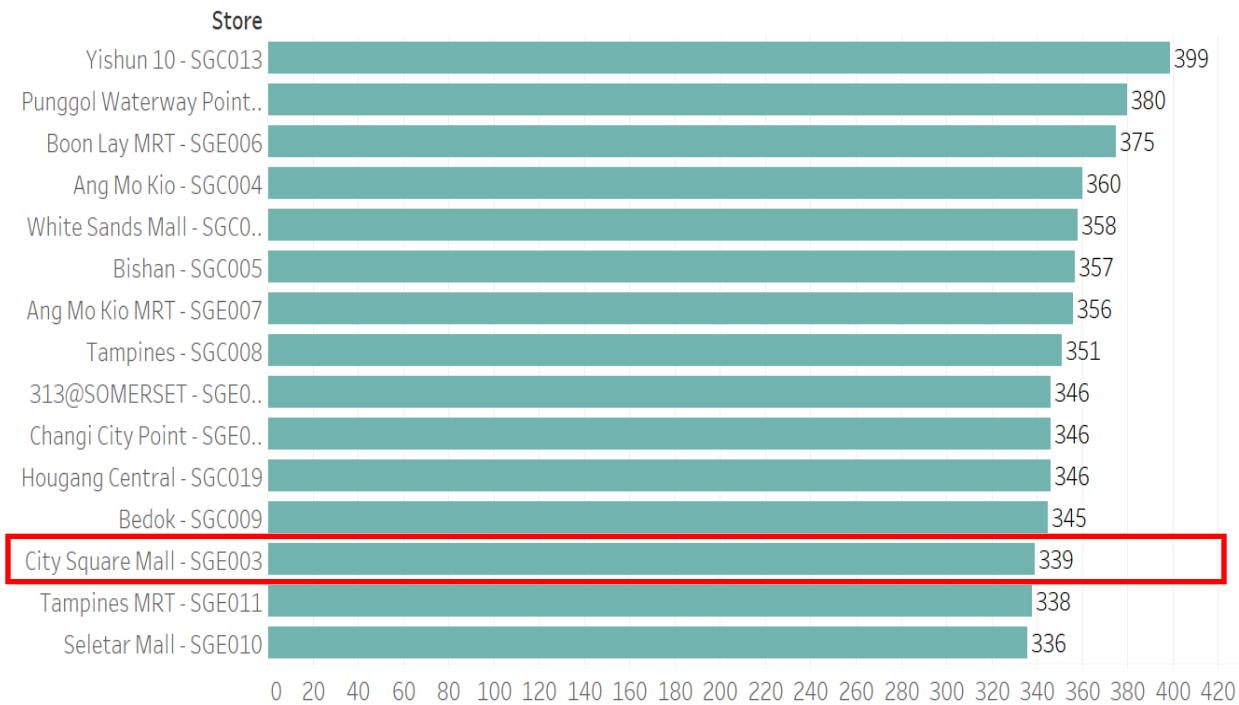


Time series analysis on the quantity ordered for City Square Mall

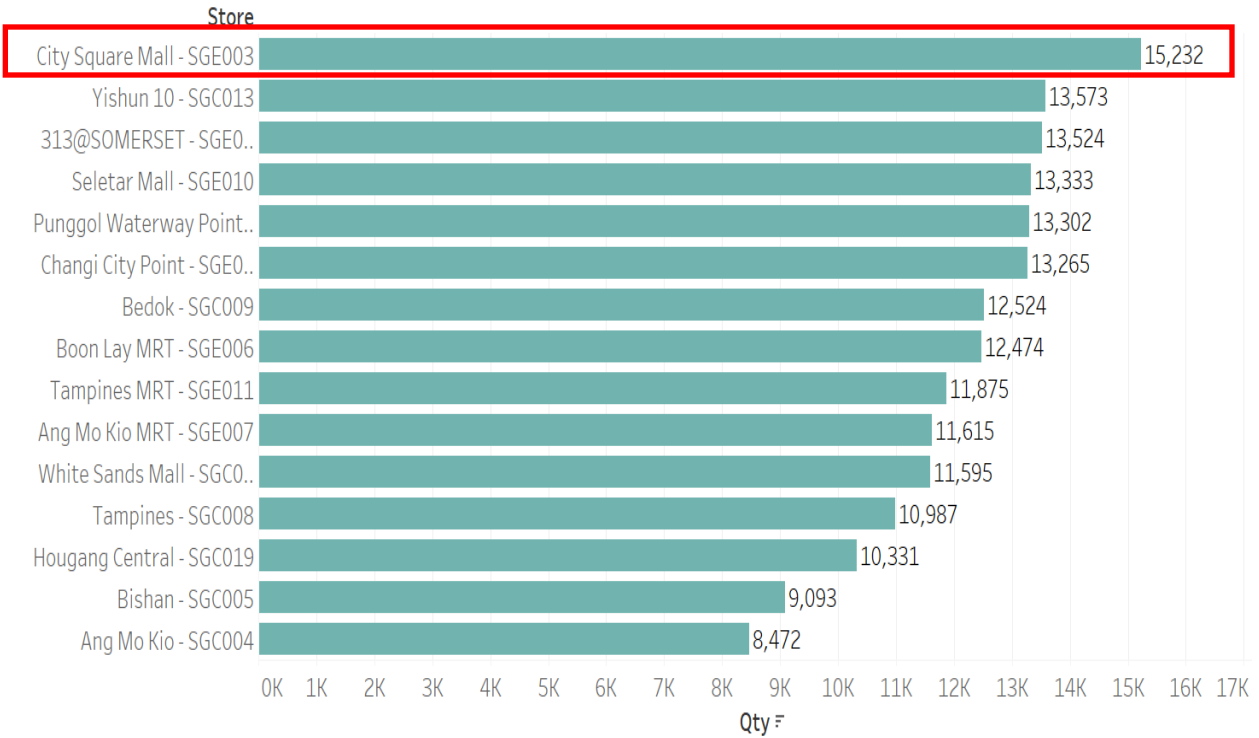


Exploratory Data Analysis

Top 15 outlets on delivery ordered

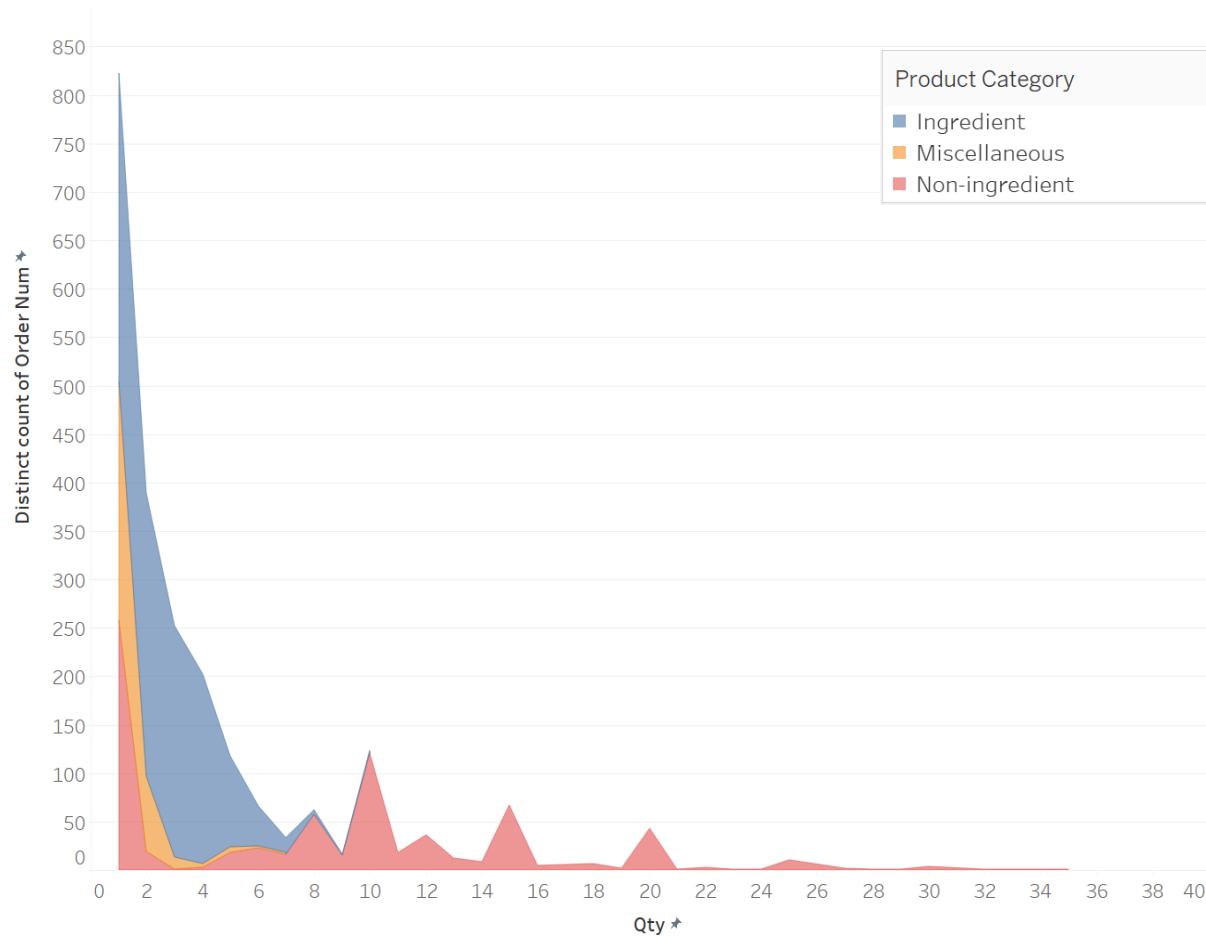


Top 15 outlets on quantity ordered



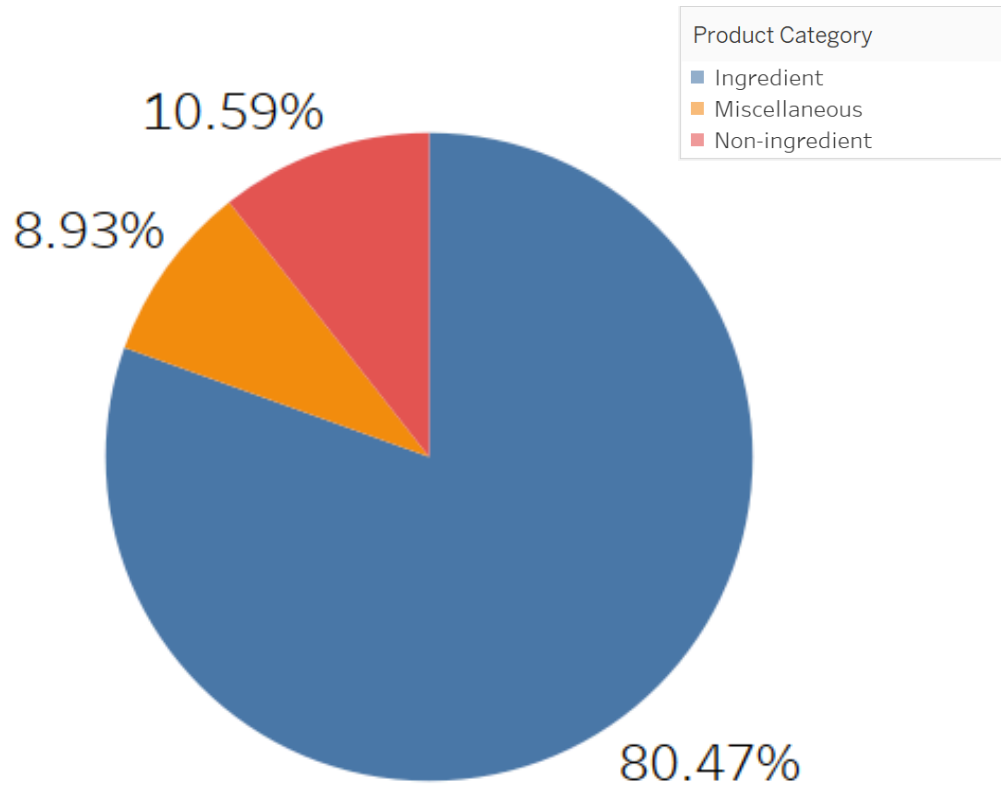
Exploratory Data Analysis

City Square Mall Quantity Ordering Trends filter by Product Category

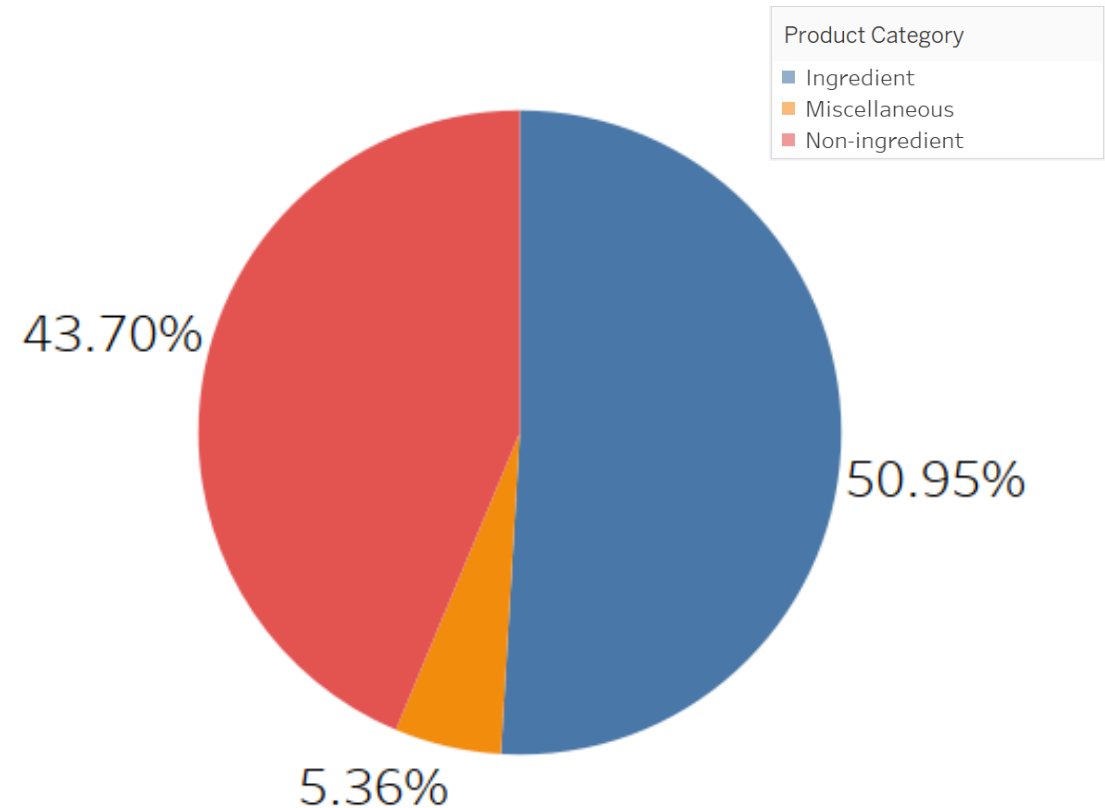


Exploratory Data Analysis

Composition of Average Quantity Ordered by Product Category

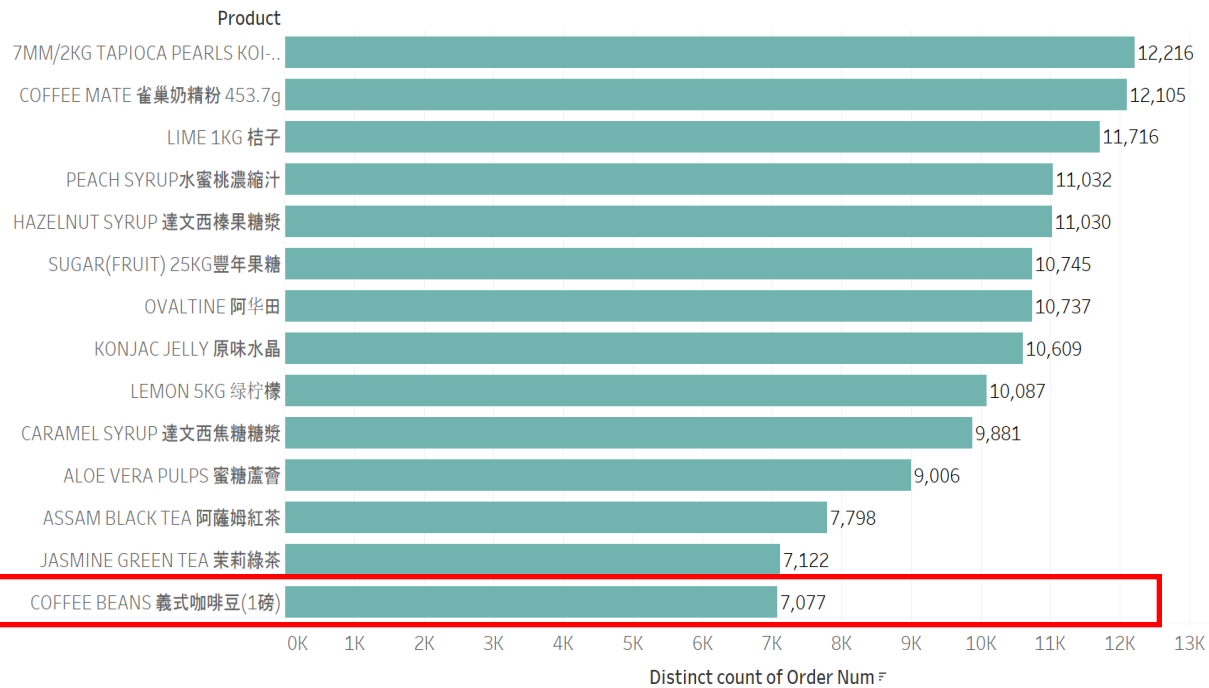


Composition of Quantity Ordered by Product Category for City Square Mall

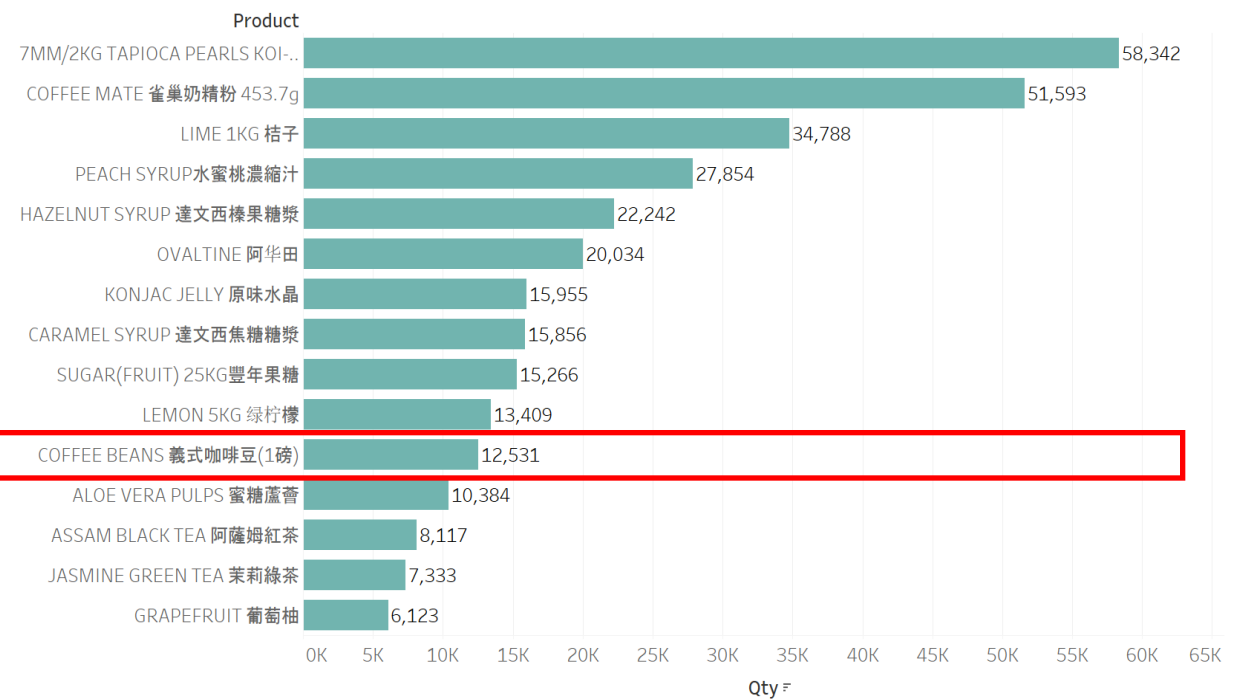


Exploratory Data Analysis

Top 15 Products on Delivery Ordered

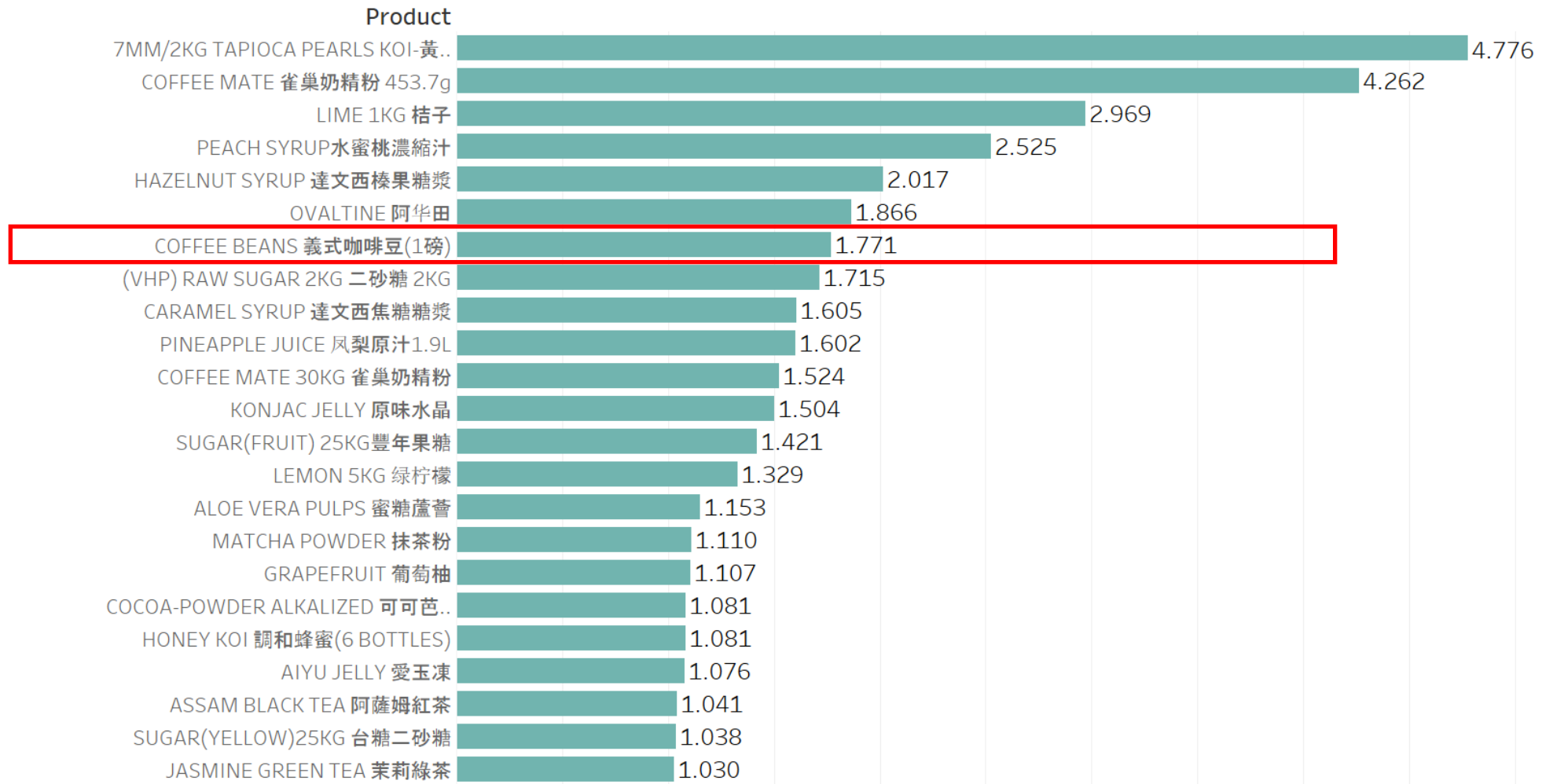


Top 15 Products on Quantity Ordered



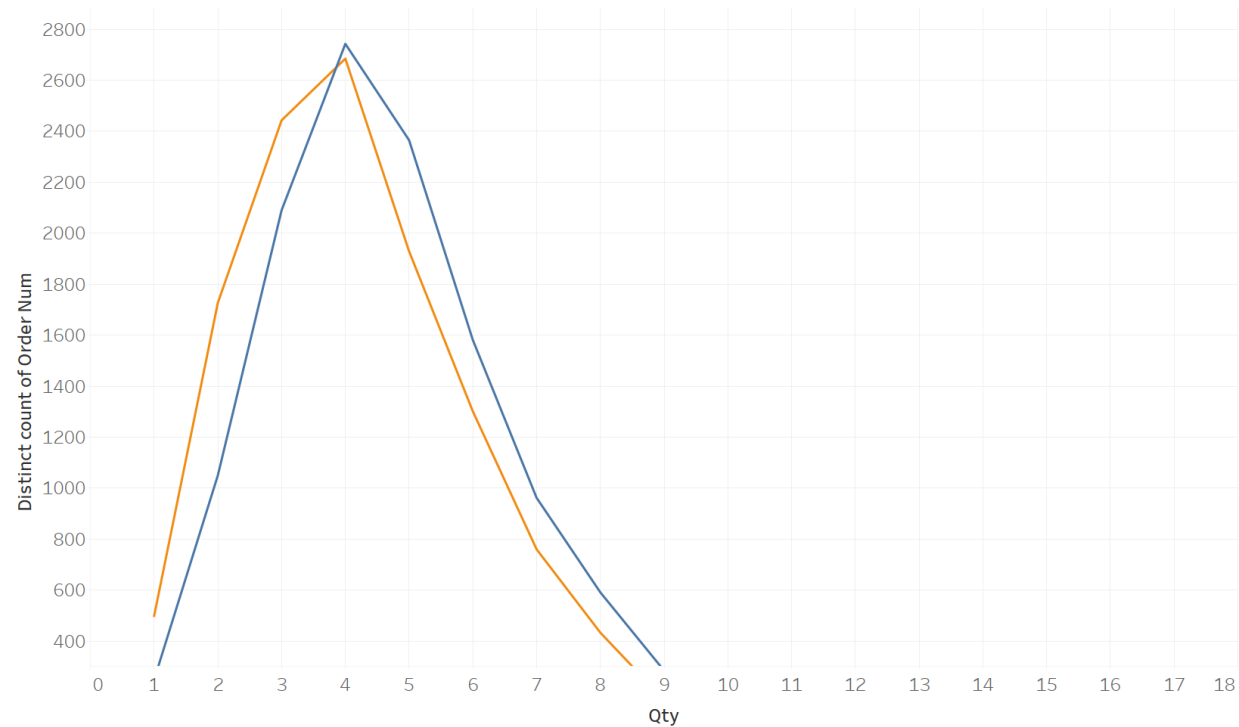
Exploratory Data Analysis

Top Products Ordered by Ratio

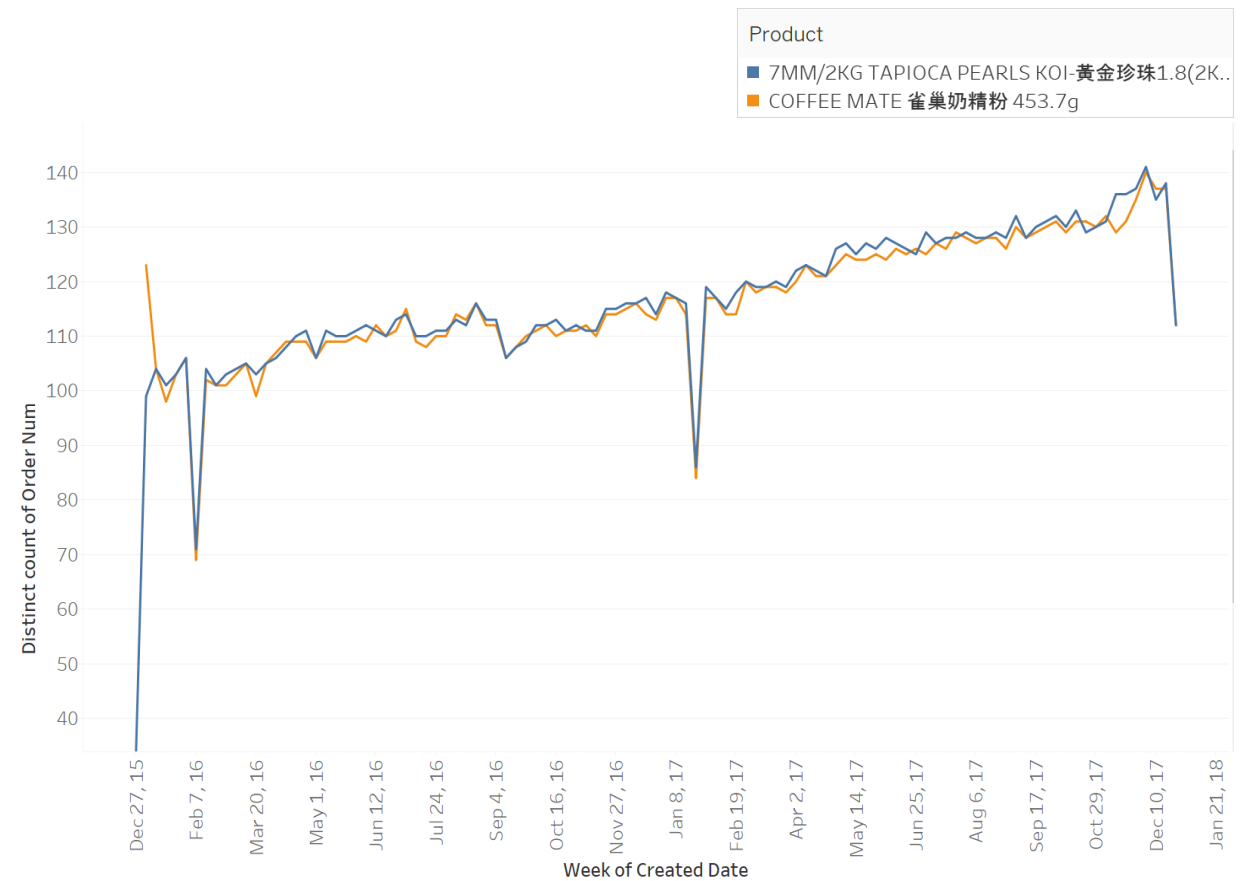


Exploratory Data Analysis

Number Of Delivery Per Quantity Ordered

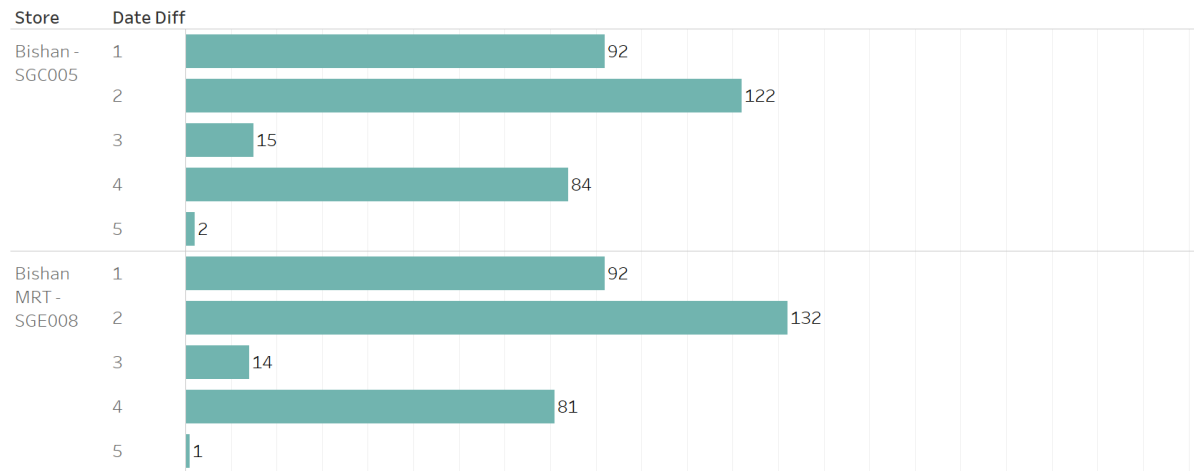


Time Series Trend of No. of Delivery Per Week

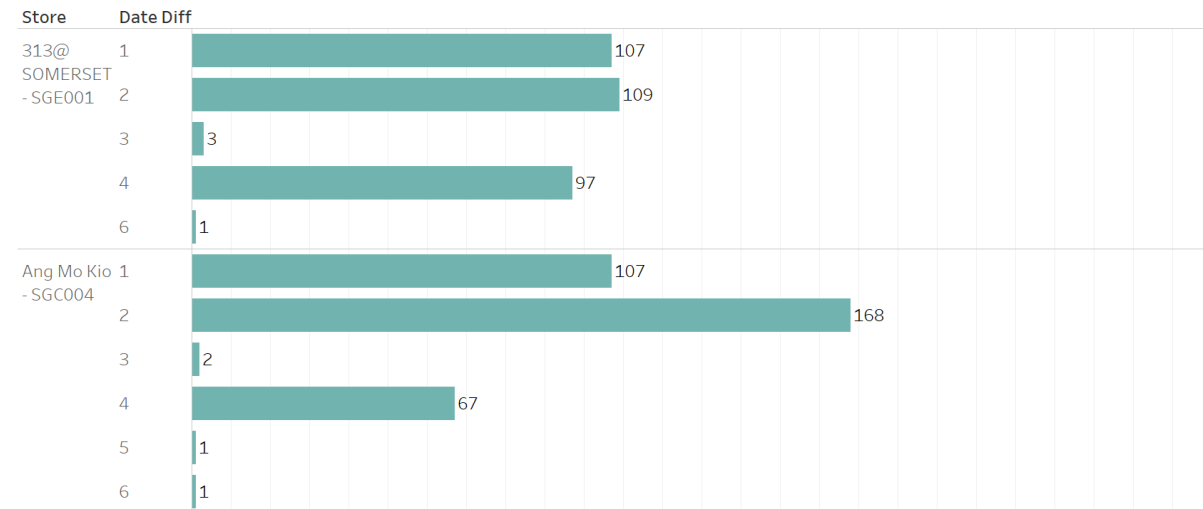


Exploratory Data Analysis

Segment 1:

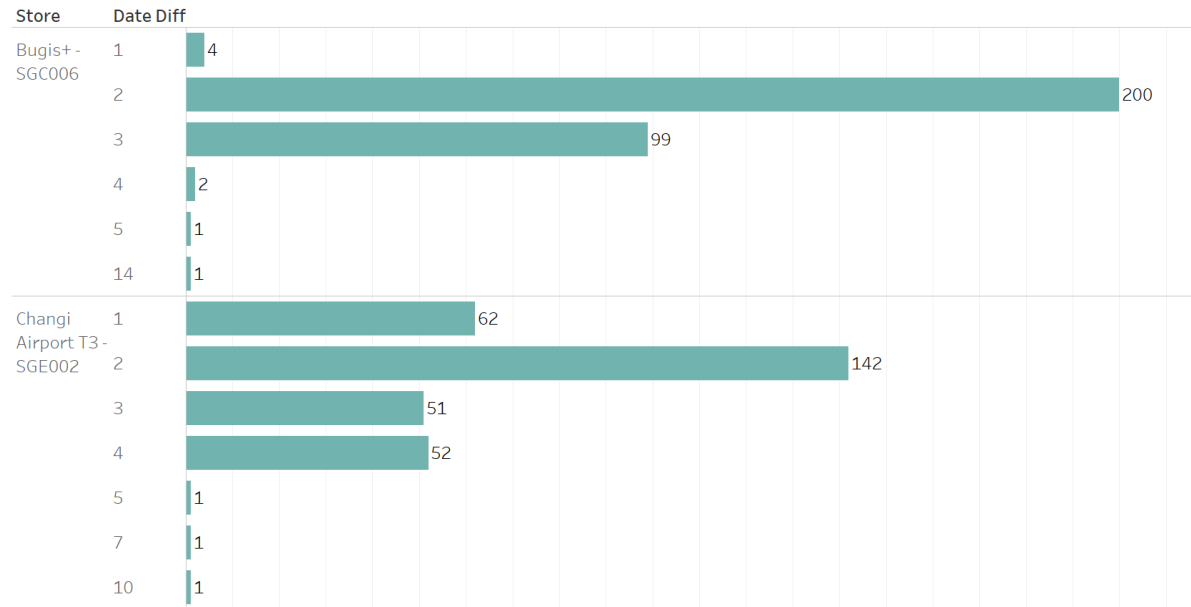


Segment 2:

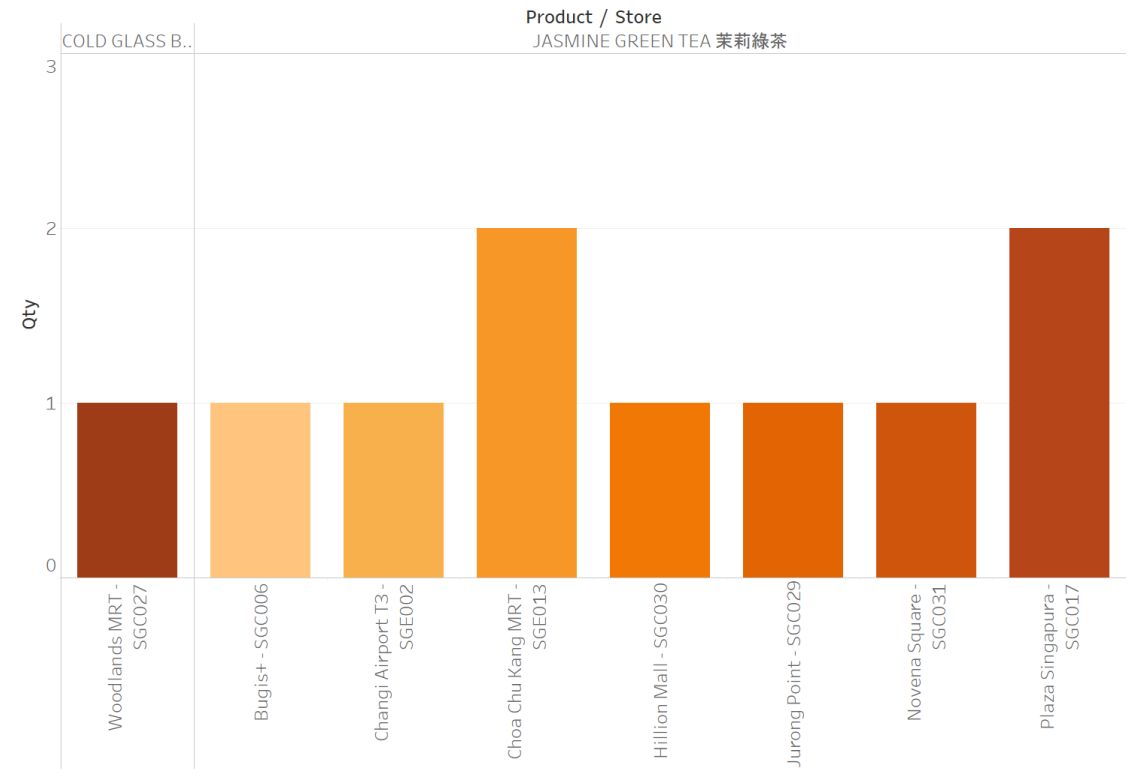


Exploratory Data Analysis

Segment 3:

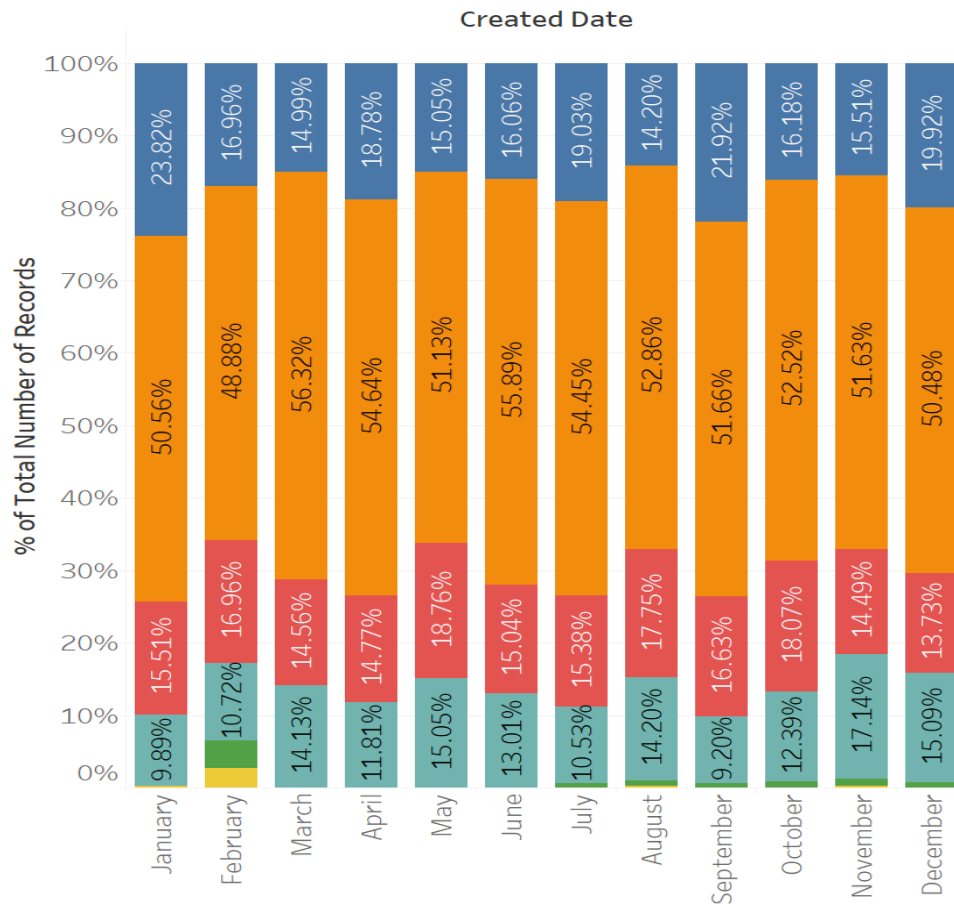


Outliers Product Composition:



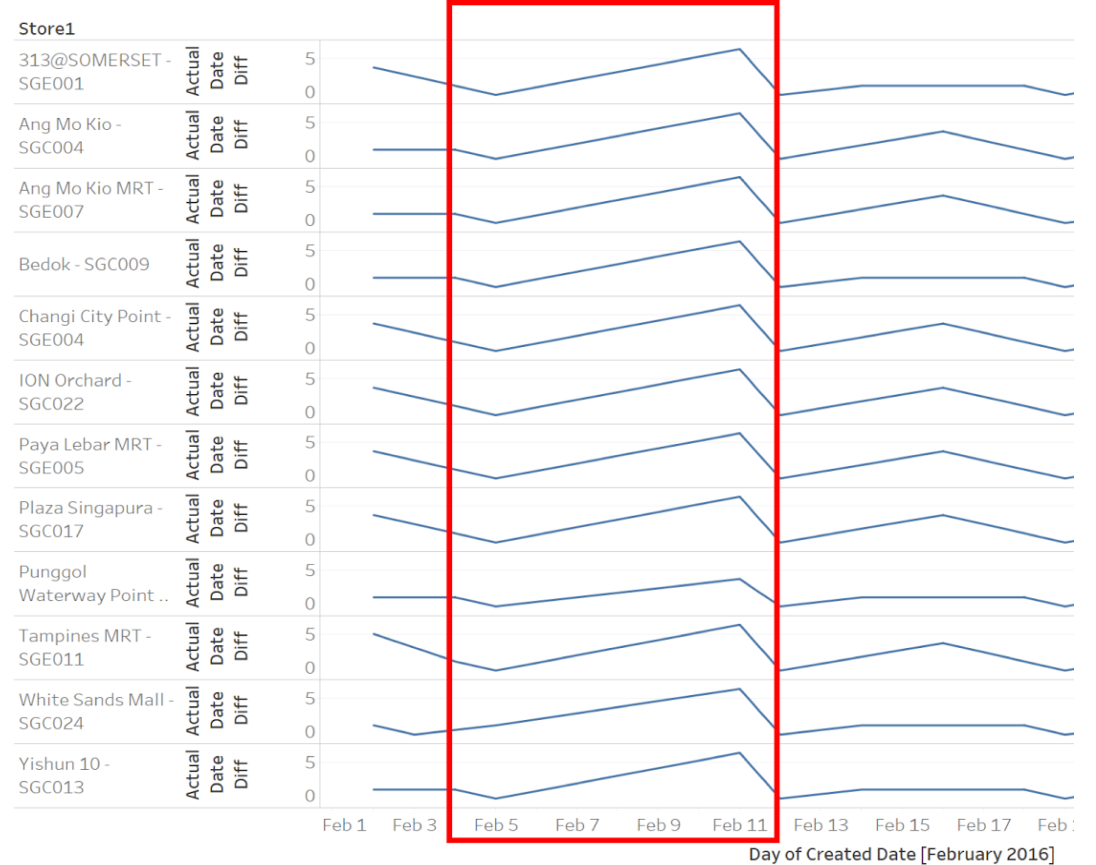
Exploratory Data Analysis

Order Frequency Composition for Monthly in 2016



Order Frequency Drilled-Down into February

Reordering Frequency (Days)



Revised Scope of Work

Completed ✓	Removed ∅	In-Progress and New Tasks +
Data cleaning on all 5 tables by identifying unrelated data to our analysis (i.e. oversea outlet delivery record which is insignificant in Singapore context)	Evaluate the effectiveness of the promotional campaigns held	Remove outliers for model building
Identify the most popular product across all outlets in year 2016 and 2017	Location analysis which is best for outlet opening based on the age group and demographics living in the area	Prepare dataset for model building
Identify the frequency of delivery across two years		Clustering for time-series similarities with wrapping method

Revised Scope of Work

Completed ✓	Removed Ø	In-Progress and New Tasks +
Identify the most popular branch across all outlets basing on the reorder quantity accumulated over 2 years		Exponential Smoothing using the weighted moving average to conduct prediction
Identify the reordering frequency in days across all outlets		Train and test the model using 70% training and 30% test model
Identify outliers of reorder frequency (days)		Monte Carlo model prediction to have another view on the optimal reorder quantity of each product

Revised Scope of Work

Completed ✓	Removed ∅	In-Progress and New Tasks +
Identify composition and reason behind the outliers in reorder frequency (days)		Comparison of the two models to evaluate which is more effective
Deep dive of outliers by months to understand which month contribute to the highest no. of outliers		Written report, final presentation, poster, insights and recommendations for KOI
Identify underlying reason behind the outliers by outlet		

Revised Scope of Work

Completed ✓	Removed ∅	In-Progress and New Tasks +
Identify time series trend of each product across different outlets		
Identify products without time series trend and label it as "No" in created field "Include?"		

Revised Methodology



CLUSTER ANALYSIS

Products that show obvious
time series patterns



MONTE CARLO

Products without obvious time
series patterns

S/N	Task		Assigned to	Week -1	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15		
				25th Dec - 31st Dec	1st Jan - 7th Jan	8th Jan - 14th Jan	15th Jan - 21st Jan	22nd Jan - 28th Jan	29th Jan - 4th Feb	5th Feb - 11th Feb	12th Feb - 18th Feb	19th Feb - 25th Feb	26th Feb - 4th Mar	5th Mar - 11th Mar	12th Mar - 18th Mar	19th Mar - 25th Mar	26th Mar - 1st Apr	2nd Apr - 8th Apr	9th Apr - 15th Apr	16th Apr - 22nd Apr		
1	Gathering Requirements	Confirm Project Sponsor	All																			
2		Gather Requirements	All																			
3	Data Gathering	Gather Data from Sponsor	Ryan																			
4	Project Proposal	Refining Project Scope	All																			
5		Proposal Preparation	Angie, Shi Jia																			
6		Wiki Page Preparation	Ryan, Shi Jia																			
7	Research	Literature Review	Angie, Shi Jia																			
8	Proposal Deadline - 14th January 2017 (Week 1)																					
9	Data Exploratory	Data Preparation	All																			
10		Data Cleaning	All																			
11		Data Exploration	All																			
12	Data Exploratory	Generate Findings	All																			
13	Project Revision	Review Progress and Findings with Sponsor	All																			
14	Interim	Interim Report Preparation	Angie, Shi Jia																			
15		Interim Presentation Preparation	Angie, Ryan																			
16		Interim Wiki Update	Ryan																			
17	Interim Practicum Presentation - 26th Feb to 4th March (Week 8)																					
18	Model Building	Data Normalisation	All																			
19		Quantitative Forecasting Analysis	All																			
20		Correlation Analysis	All																			
21		Model Validation	All																			
22	Insights and Recommendations	Generate Visualizations from Analysis Results	All																			
23		Generating Insights	All																			
24		Formulate Recommendations	All																			
25	Project Review	Review Progress and Findings with Sponsor	All																			
26		Align Final Deliverables with Sponsor Requirement	All																			
27	Final Preparation	Abstract Paper Preparation	Shi Jia																			
28		Full Paper Preparation	Shi Jia																			
29		Final Wiki Update	Ryan																			
30		Final Presentation Preparation	Angie																			
31		Final Paper Preparation	Angie, Ryan																			
32		Poster Submission Preparation	Shi Jia, Ryan																			
33	Abstract Paper Submission - 1st April 2018 (Week 12)																					
34	Full Paper Submission - 8th April 2018 (Week 13)																					
35	Undergraduate Conference on Data Analytics Conference Day - 14th April to 15th April (Week 14)																					
36	Analytics Practicum Final Submission - 22nd April 2017 (Week 15)																					

Q & A

