Insights Analyst - Technical Exercise

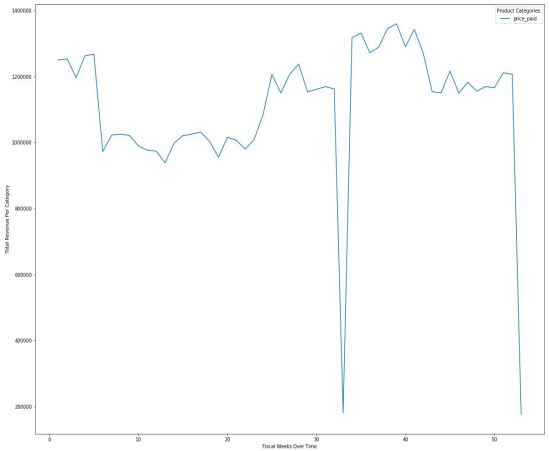
Amy Vong

Confidential

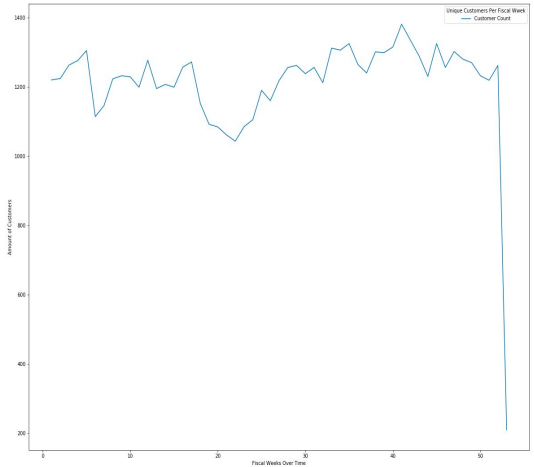
Question 1

Use a time series plot to: ***BY FISCAL WEEK***

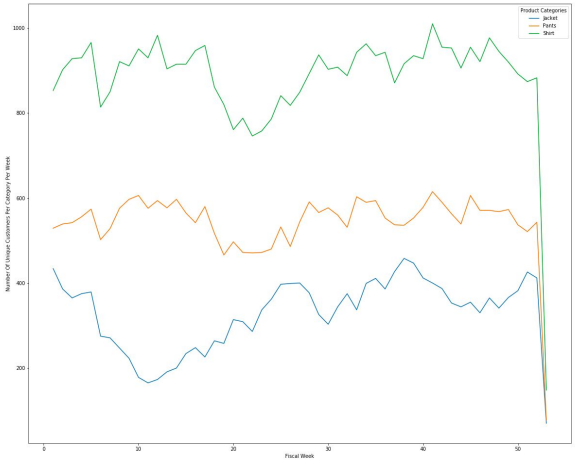
1)Total spend per week across the customer base over time



2) **Number of unique customers per week** across the customer base over time.



3) Number of unique customers, by category, per week across the customer base over time, broken out by category (Shirt / Jacket / Pants) to compare purchasing

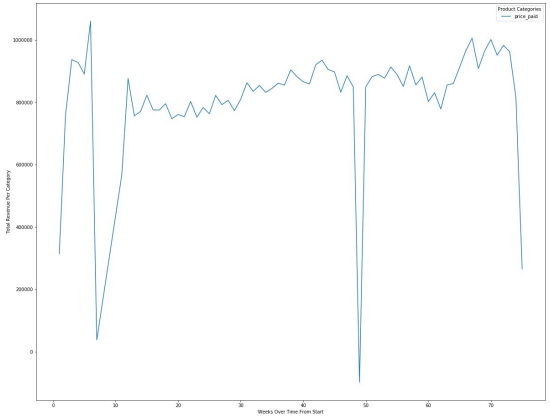


Confidential

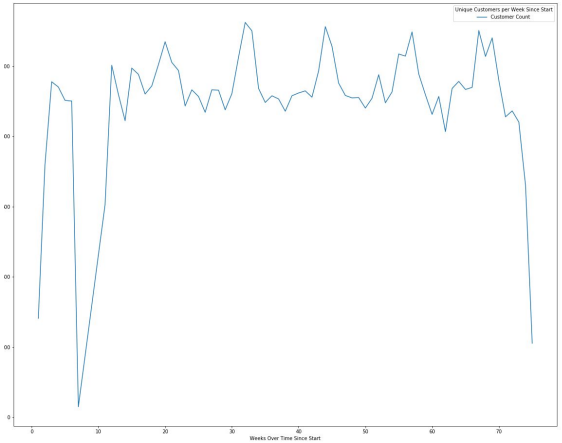
Question 1

Use a time series plot to: ***BY BEGINNING OF DATASET***

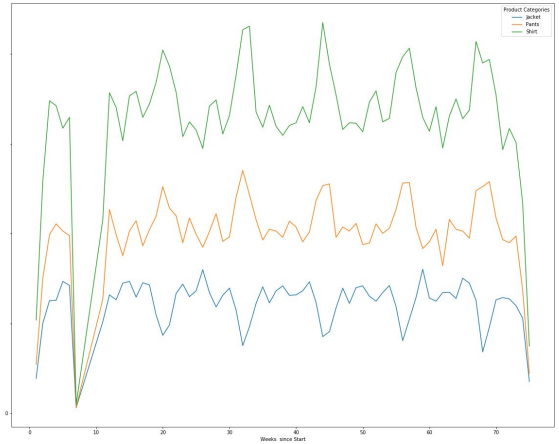
1)Total spend per week across the customer base over time



2) **Number of unique customers per week** across the customer base over time.



3) Number of unique customers, by category, per week across the customer base over time, broken out by category (Shirt / Jacket / Pants) to compare purchasing

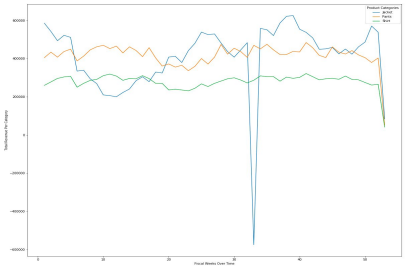


Confidential

Question 2

There are at least two significant issues in the data which could affect analysis. Please comment on what these are, and then make any fixes as appropriate before continuing, justifying your assumptions.

1) There was an anomaly in 2018 with a Jacket transaction that caused a substantial loss in Revenue. I located that anomaly and found that it was a one-off transaction. Because of the volume of this single unit, it skewed the data. The line can be removed from the dataframe to better analyze trends around product and revenue, but with the caveat that the client knows that it was one of the biggest impacts on that year and that fiscal week.

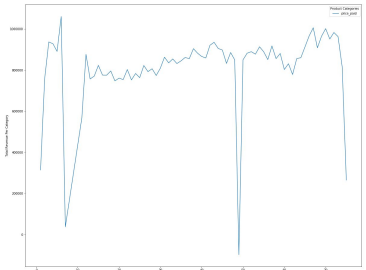


```
df2018=df[(df['order_date_year'] == 2018)]
df2018.sort_values(by='price_paid', ascending=True)
```

	customer_id	order_id	order_date	price_paid	product_category	Start of Data	weeks_since_start	order_date_week	order_date_mo
83468	494ca220e3920d02cdc5	ae23b0c6835efa748fd5	2018-08-14	-1000000.000000	Jacket	2014-08-16	209.0	33	

2) Data is missing from February 2015 - 2015. This causes a big discrepancy when trying to compare revenue, product, and customer sales behavior. In order to accurate measure growth over time, comparisons should be made from 2016 in relation to P&L for full breadth and depth.

	order_date_year	order_date_month	price_paid
5	2015	1	1.065770e+06
6	2015	6	5.420924e+05
7	2015	7	8.709538e+05
8	2015	8	7.884884e+05
9	2015	9	7.708793e+05
10	2015	10	8.230563e+05
11	2015	11	7.757477e+05
12	2015	12	7.754040e+05



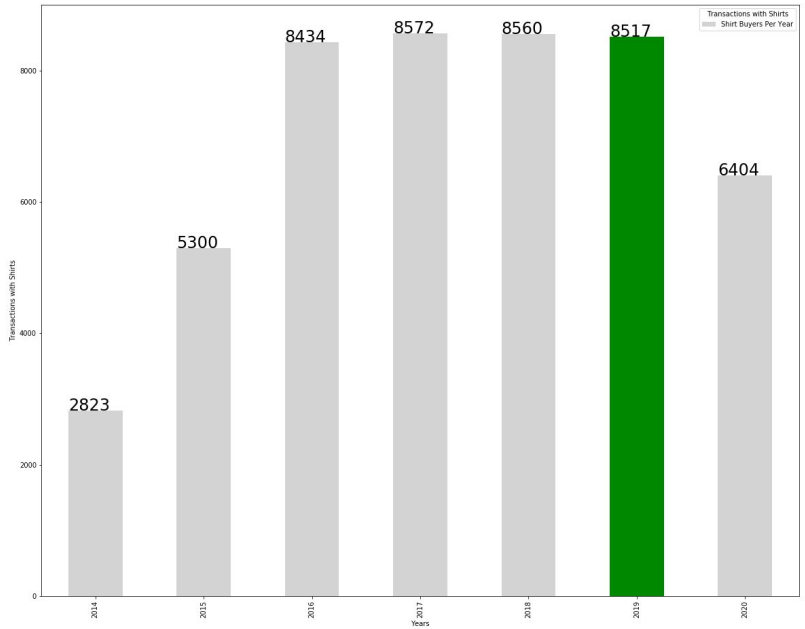
Confidential

Question 3

According to the data, how many Shirt buyers were there in 2019? I.e. customers who had a shirt in an order in 2019?

In 2019, **8517** transactions had at least one Shirt item in their purchase.

	order_date_year	Number of Trans with Shirts
0	2014	2823
1	2015	5300
2	2016	8434
3	2017	8572
4	2018	8560
5	2019	8517
6	2020	6404



Question 4

Find customers who meet BOTH the following criteria ('2019 First-Time Shirt Buyers'):

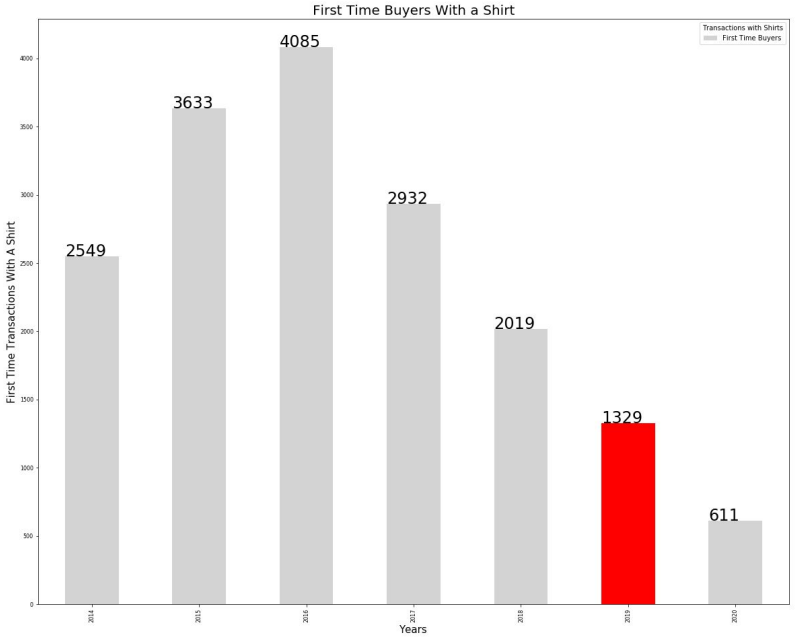
- *purchased for the first time in 2019, AND had a Shirt in their first order (you can use order_date to find the first order).*

In 2019:

8517 orders had at least one Shirt item in that purchase.

1329 of those customers were buying for the very first time.

	order_date_year	Number Of Trans With Shirts	First Time Shirt Buyers	% of New Buyers
0	2014	2823	2549	90.0%
1	2015	5300	3633	69.0%
2	2016	8434	4085	48.0%
3	2017	8572	2932	34.0%
4	2018	8560	2019	24.0%
5	2019	8517	1329	16.0%
6	2020	6404	611	10.0%



Confidential

Question 4

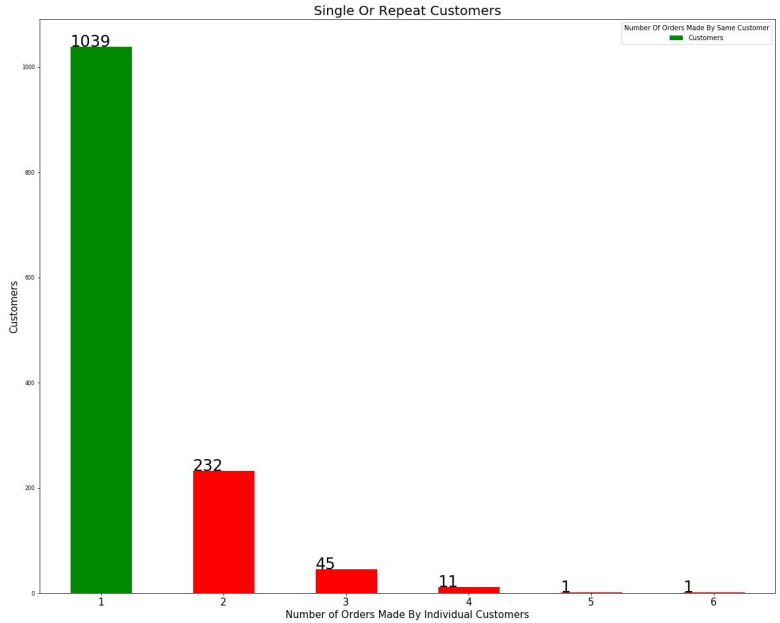
Then, using this set of customers ('2019 First-Time Shirt Buyers'):

- *how many customers have only ever made one order (one-time-only buyers)?*
- *how many customers went on to make at least one more order (repeat buyers)?*

Of the **1329** customers that were 2019 First-Time Shirt Buyers

1039 customers only ever made one single order from 2019-2020 (end of the data set).

250 customers went on to make at least one more order.



Question 4

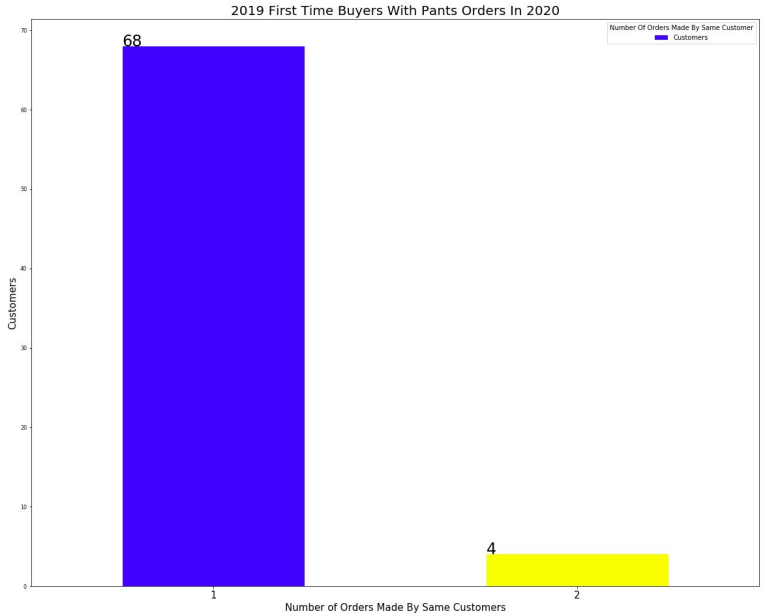
Then, using this set of customers ('2019 First-Time Shirt Buyers'):
- *how many customers went on to buy Pants in 2020?*

Of the **250** additional sales made by the 2019 First-Time Shirt Buyers in 2020

74 went on to buy Pants in 2020

68 customers only ever made one order containing Pants in 2020

4 customers went on to make at least one more order of Pants in 2020.



Amy Vong
Lexer 2021

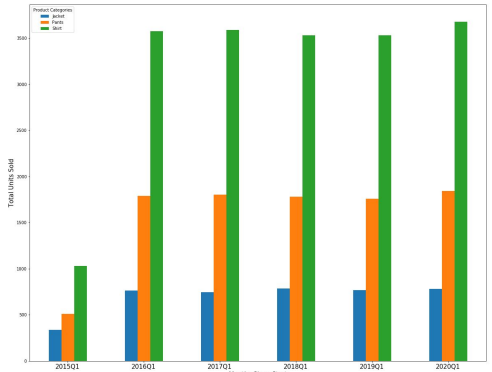
Confidential

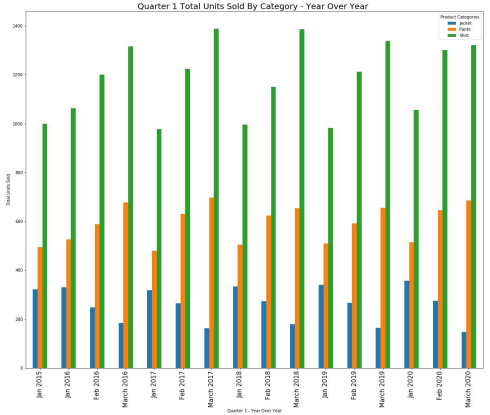
Question 5

The client you are working with asks, "We're trying to develop our marketing strategy for January–March next year. We're not sure whether to promote Jackets, Pants, Shirts, or some combination of these. Based on the data, especially any seasonal patterns, what would you recommend?"

Taking all 3 product categories and pulling data for Quarter 1 year over year, the data to the right shows how each category contributed with total units sold and total sales revenue.

1) Shirts contributes the lowest in revenue from the category groups in Quarter 1 every year. Shirts consistently drive 23%- 27% of revenue contribution.
2) Jackets contributes the highest in revenue from the category groups in Quarter 1 every year. Shirts consistently drive 31%- 42% of revenue contribution.
3) Pants contributes in the middle in revenue from the category groups in Quarter 1 every year. Pants drive 35%- 41% of revenue contribution. The category has contributed 30% of the units sold and a 40% contribution to revenue. The Average Order Value aligns with the average total purchase allowing an assumption that the price point for the average customer is accessible.





	quarter	product_category	Percent to Total Revenue	Percent to Total Units Sold
0	2015Q1	Jacket	42.0%	18.0%
1	2015Q1	Pants	35.0%	27.0%
2	2015Q1	Shirt	23.0%	55.0%
3	2016Q1	Jacket	33.0%	12.0%
4	2016Q1	Pants	40.0%	29.0%
5	2016Q1	Shirt	27.0%	58.0%
6	2017Q1	Jacket	31.0%	12.0%
7	2017Q1	Pants	41.0%	29.0%
8	2017Q1	Shirt	27.0%	58.0%
9	2018Q1	Jacket	34.0%	13.0%
10	2018Q1	Pants	40.0%	29.0%
11	2018Q1	Shirt	27.0%	58.0%
12	2019Q1	Jacket	33.0%	13.0%
13	2019Q1	Pants	40.0%	29.0%
14	2019Q1	Shirt	27.0%	58.0%
15	2020Q1	Jacket	32.0%	12.0%
16	2020Q1	Pants	41.0%	29.0%
17	2020Q1	Shirt	27.0%	58.0%

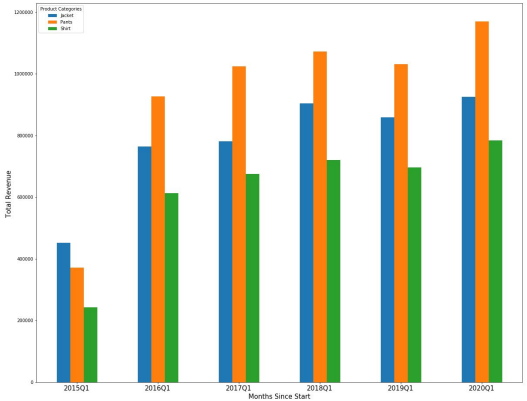
Confidential

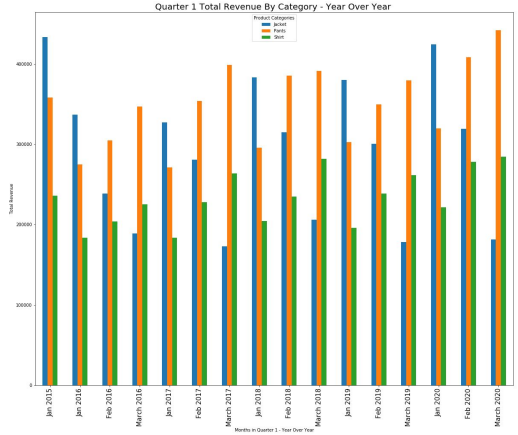
Question 5

The client you are working with asks, "We're trying to develop our marketing strategy for January–March next year. We're not sure whether to promote Jackets, Pants, Shirts, or some combination of these. Based on the data, especially any seasonal patterns, what would you recommend?"

Taking all 3 product categories and pulling data for Quarter 1 year over year, the data to the right shows how each category contributed by total revenue/

1) Shirts contribute to a majority of the total unit sold in Quarter 1 every year. With a lower Average Order Value and high units sold in customer orders, it has contributed 55-58%% in all category sales, but only 23%- 27% of revenue contribution.
2) Jackets has the highest Average Order Value and the least units sold at 12%-18% of overall units sold. However, the category contributes 31%- 42% of Total Revenue
3) Pants is in the middle contributing to 27%-29% of the units sold and 35%- 41% contribution to revenue.





	quarter	product_category	Percent to Total Revenue	Percent to Total Units Sold
0	2015Q1	Jacket	42.0%	18.0%
1	2015Q1	Pants	35.0%	27.0%
2	2015Q1	Shirt	23.0%	55.0%
3	2016Q1	Jacket	33.0%	12.0%
4	2016Q1	Pants	40.0%	29.0%
5	2016Q1	Shirt	27.0%	58.0%
6	2017Q1	Jacket	31.0%	12.0%
7	2017Q1	Pants	41.0%	29.0%
8	2017Q1	Shirt	27.0%	58.0%
9	2018Q1	Jacket	34.0%	13.0%
10	2018Q1	Pants	40.0%	29.0%
11	2018Q1	Shirt	27.0%	58.0%
12	2019Q1	Jacket	33.0%	13.0%
13	2019Q1	Pants	40.0%	29.0%
14	2019Q1	Shirt	27.0%	58.0%
15	2020Q1	Jacket	32.0%	12.0%
16	2020Q1	Pants	41.0%	29.0%
17	2020Q1	Shirt	27.0%	58.0%

Question 5

The client you are working with asks, "We're trying to develop our marketing strategy for January–March next year. We're not sure whether to promote Jackets, Pants, Shirts, or some combination of these. Based on the data, especially any seasonal patterns, what would you recommend?"

Taking a look at Total Units Sold and Total Revenue for Q1 over time:

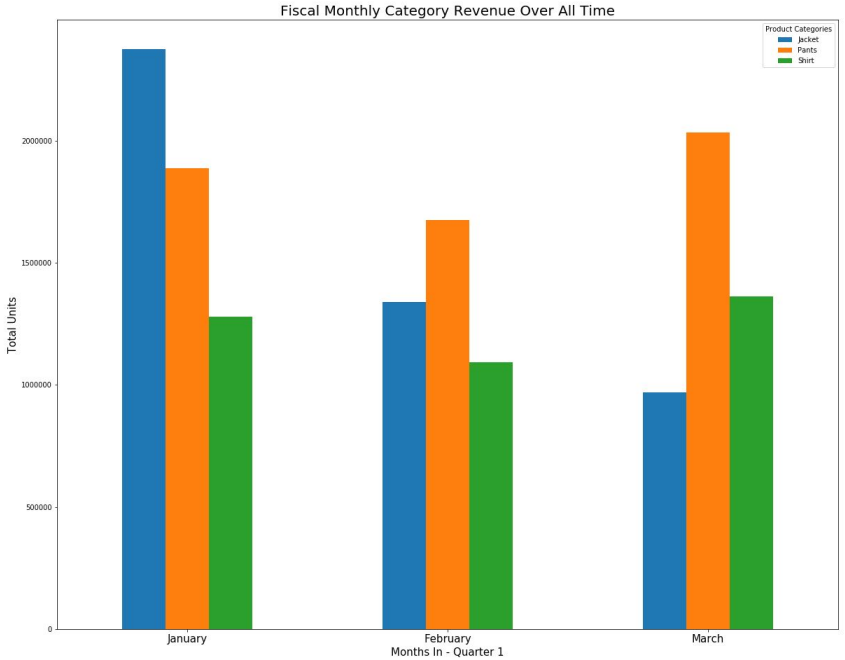
Two ways to approach marketing in these months:

1) Market the Most Profitable categories during their best months to convert better sales and higher units sold in top performing categories during their peak time.
2) Market the Least Profitable categories during their months of opportunity to potentially convert sales into new categories from customers who are purchasing other products.

I. Shirts
   A. March - Most Profitable
   B. January - Least Profitable

II. Jackets
   A. January - Most Profitable
   B. March - Least Profitable

III. Pants
   A. February and March - Most Profitable
   B. January - Profitable
      1. *Pants has typically outperformed at least one other category Month over Month. It would be a great item to combine for marketing with either of the other categories.*



Confidential