UNIVERSITY OF TECHNOLOGY SYDNEY

Studio 2: Specialisation (Individual)

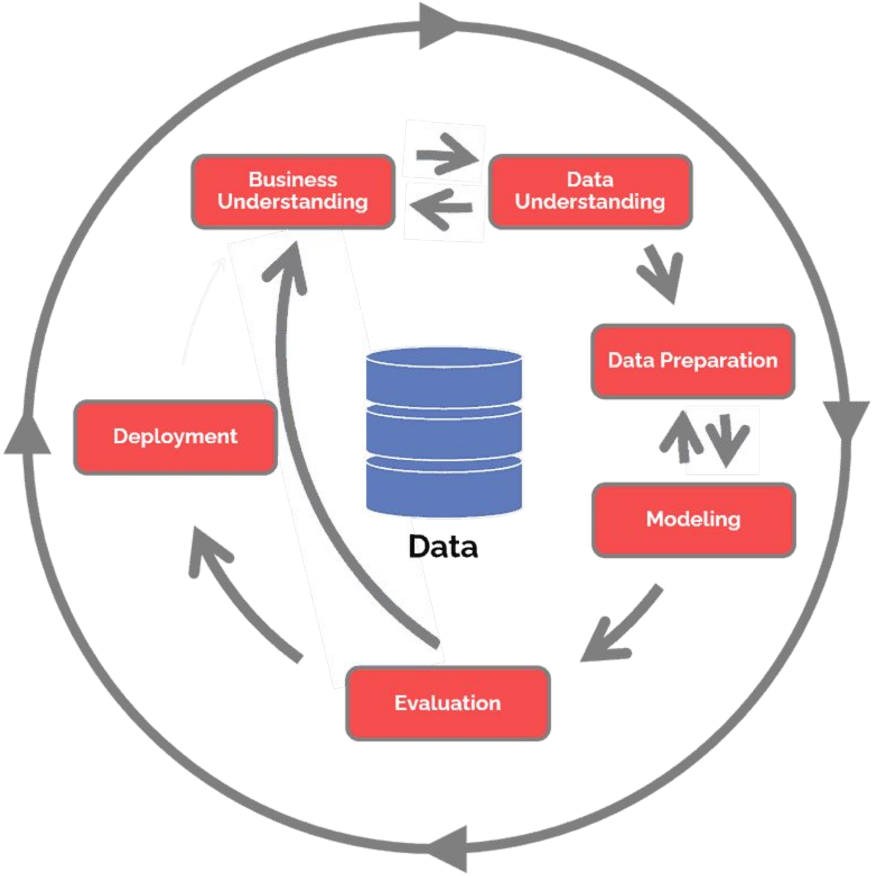
Bibek Bhandari - 14338471

Table of Contents

Executive Summary

The CRISP-DM (Cross-Industry Standard Process for Data Mining) model is a widely adopted framework for conducting data analysis projects. It consists of six iterative steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

This executive summary highlights the process of using the CRISP-DM model in Project 2 - Identifying means and ways to increase the revenue and customer satisfaction, including its systematic approach to problem-solving, clear project structure, and ability to uncover valuable insights from data. By following the CRISP-DM model, organizations can improve decision-making, optimize processes, and drive business growth through effective data-driven strategies. It provides a comprehensive methodology that enables the analysts to navigate complex data analysis projects efficiently, ensuring that valuable information is extracted and utilized to drive actionable outcomes.



CRISP-DM MODEL

Cross-Industry Standard Process for Data Mining(CRISP-DM) model is a framework that is used widely among organizations for guided data mining and analytics projects.

It offers a well defined structured flow consisting of six phases:

1. Business Understanding - Gain clear understanding of the business objective.
2. Data Understanding - Explore the available data and insights.
3. Data Preparation - Cleaning, integrating, transforming data for doing the analysis
4. Modelling - Select and apply proper models based on the prepared data
5. Evaluation - Access the quality of the progress and make suggestions based on the obtained results.
6. Deployment - Incorporating the models into business operations and systems to drive actionable insights and decisions.

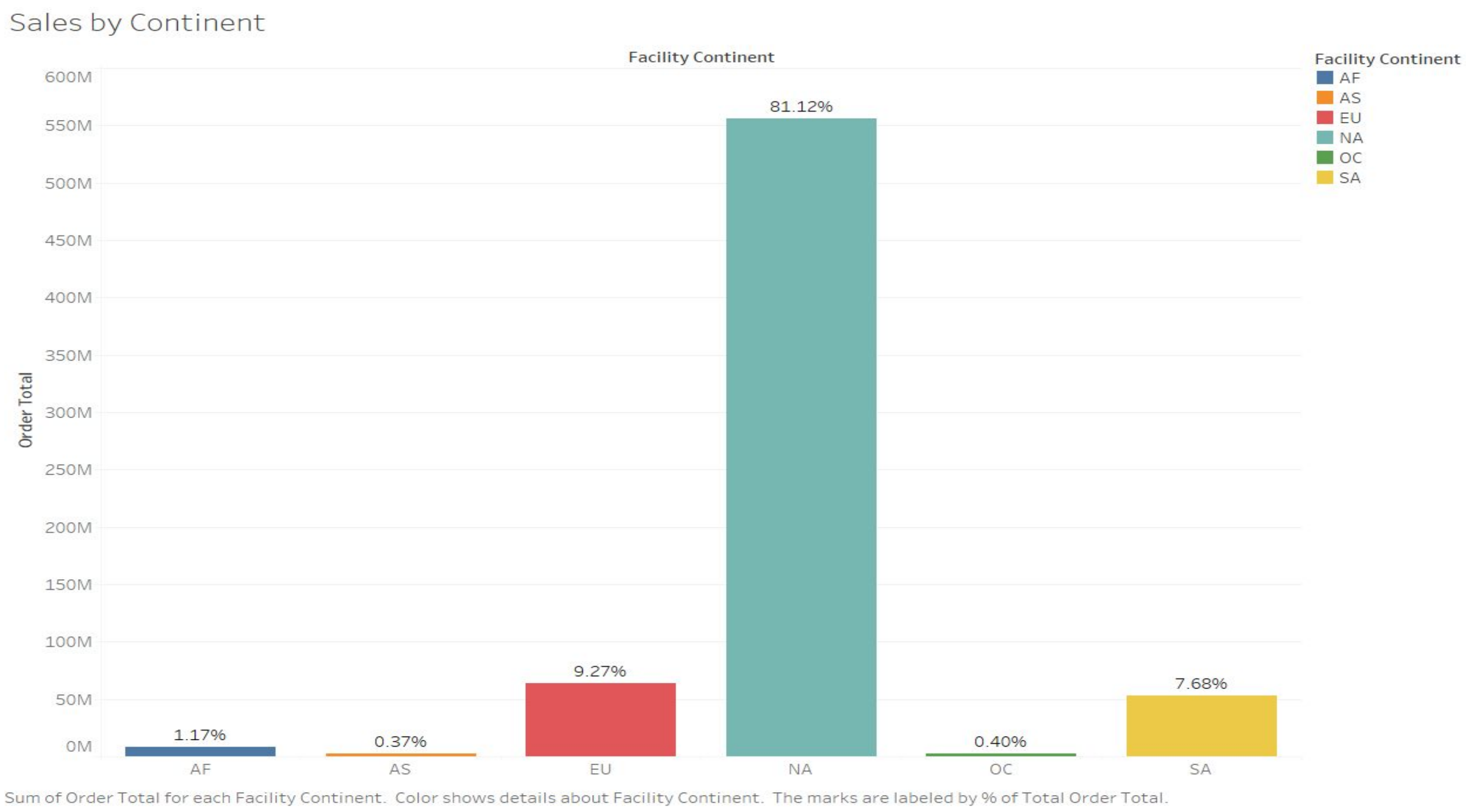
For the given Dataset we will use the model to increase the company revenue and customer satisfaction

The Process:

- Understand the Business and the Data
- Segment the data that is required for the objective i.e the revenue, profit and the customer satisfaction
- Analyze the Product Brand, Product Line, Vendor Category, Sales Rep etc, their trend and estimate related to the revenue, profit and customer satisfaction
- Analyze the top and bottom performing state(revenue)
- Provide Recommendation and Deployment Tactics

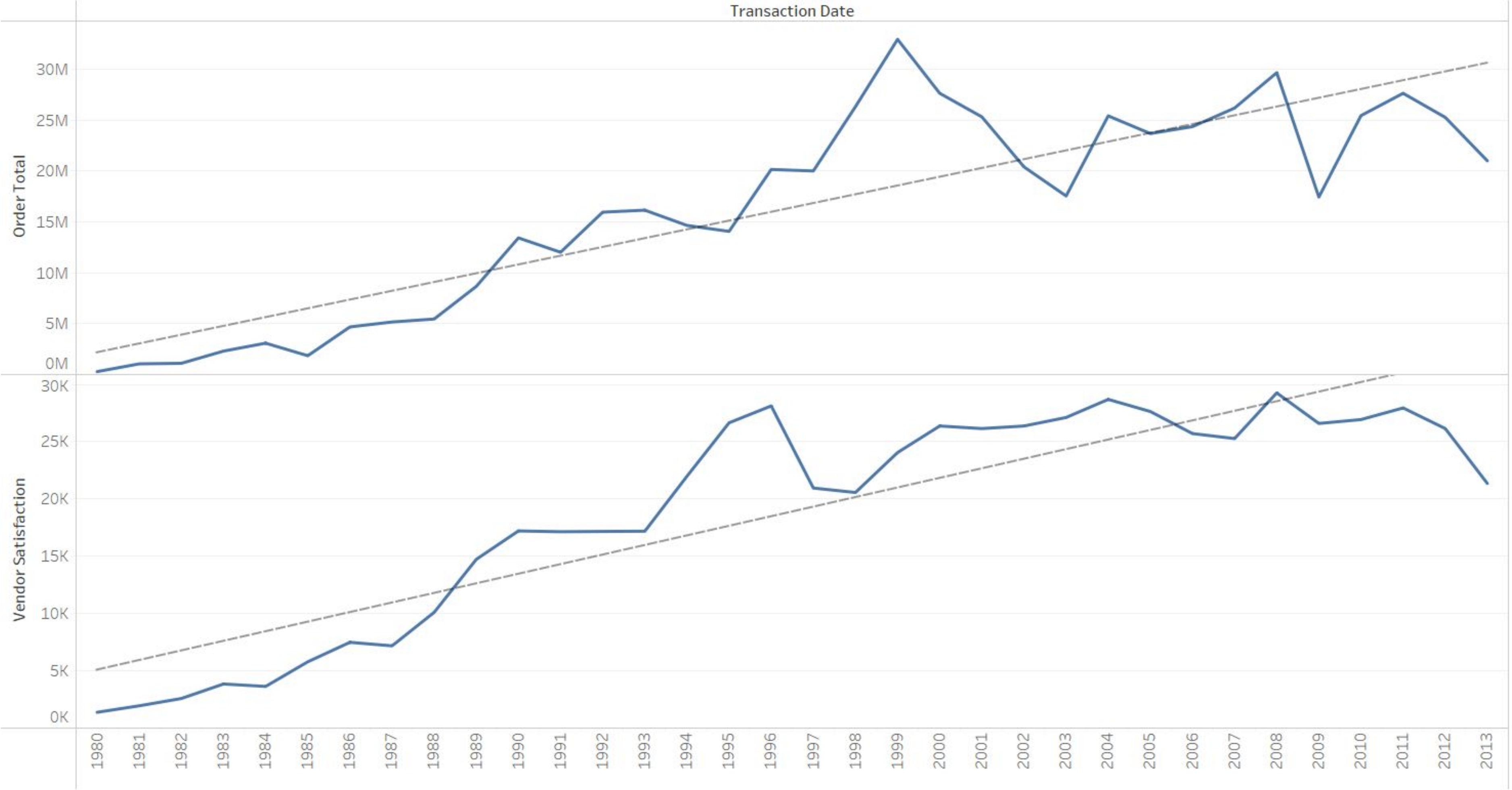
Business Understanding and Objective

- The company is a Global based US toys company which sells its product across all the seven different continents through different vendors.
- It main operation is in North America and sells mostly 2 Product Brand i.e Toys and Novelty.
- It tries to maintain an average vendor satisfaction of around 57% using different loyalty programs.
- The main objective of the company is to increase the revenue and vendor satisfaction.



The North America accounts for over 81% of the sales.

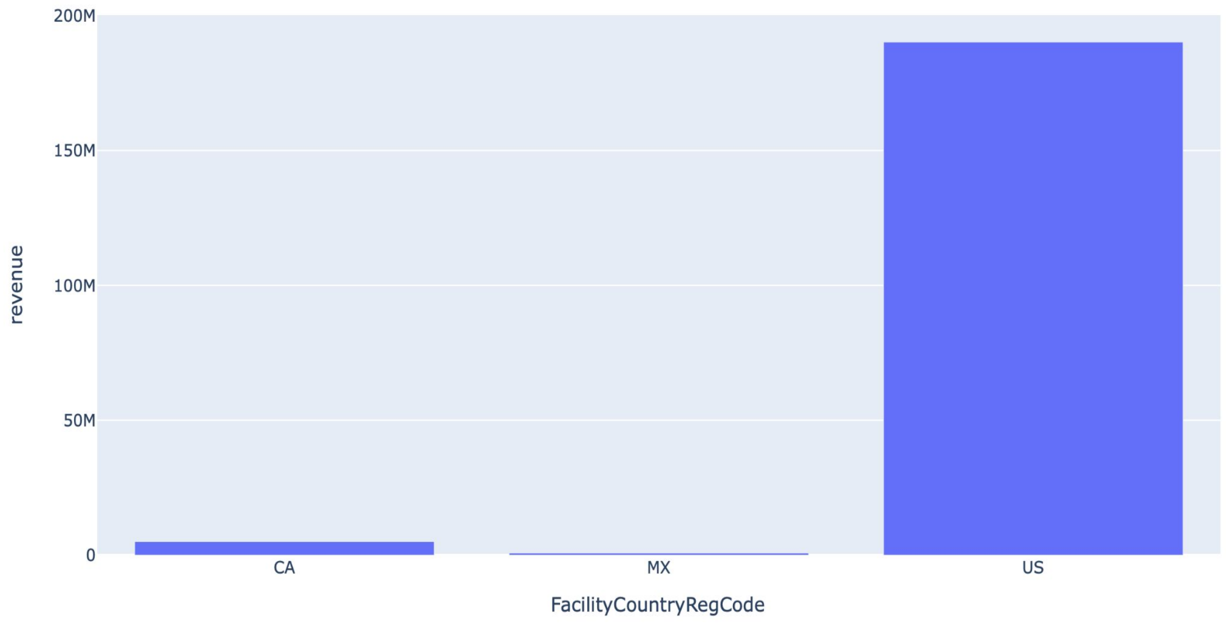
Business Revenue and Vendor Satisfaction(Time Series)



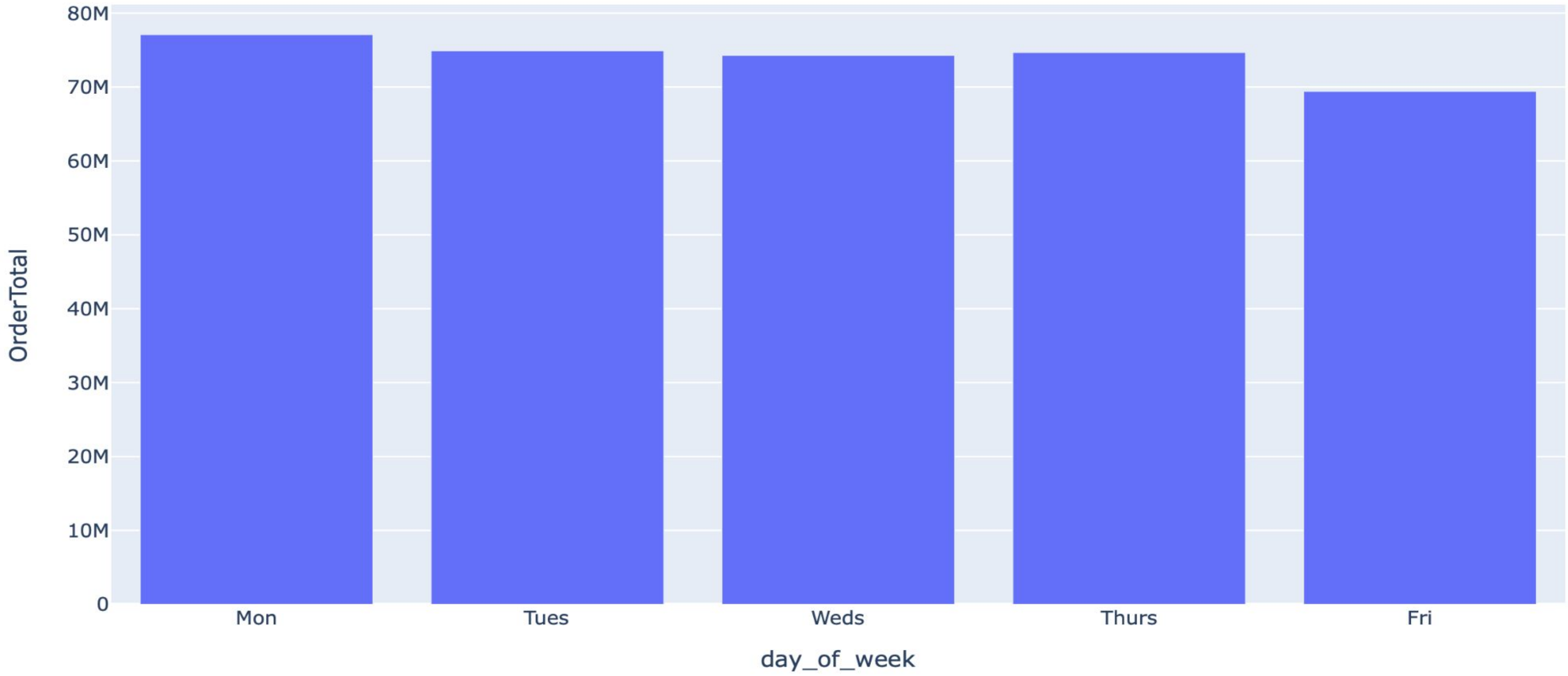
The trends of sum of Order Total and sum of Vendor Satisfaction for Transaction Date Year.

Revenue for NA countries

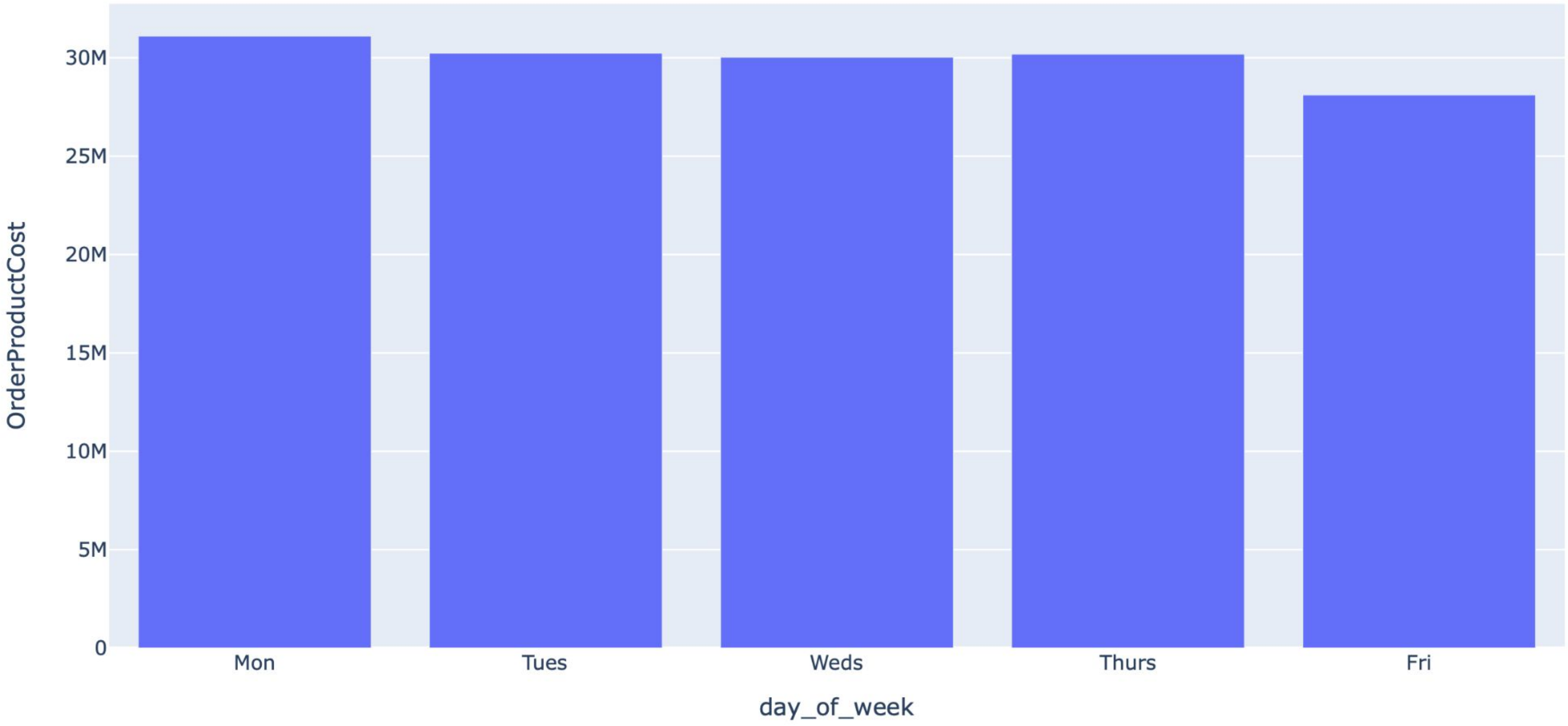
Among all the NA countries , US has the Highest revenue.



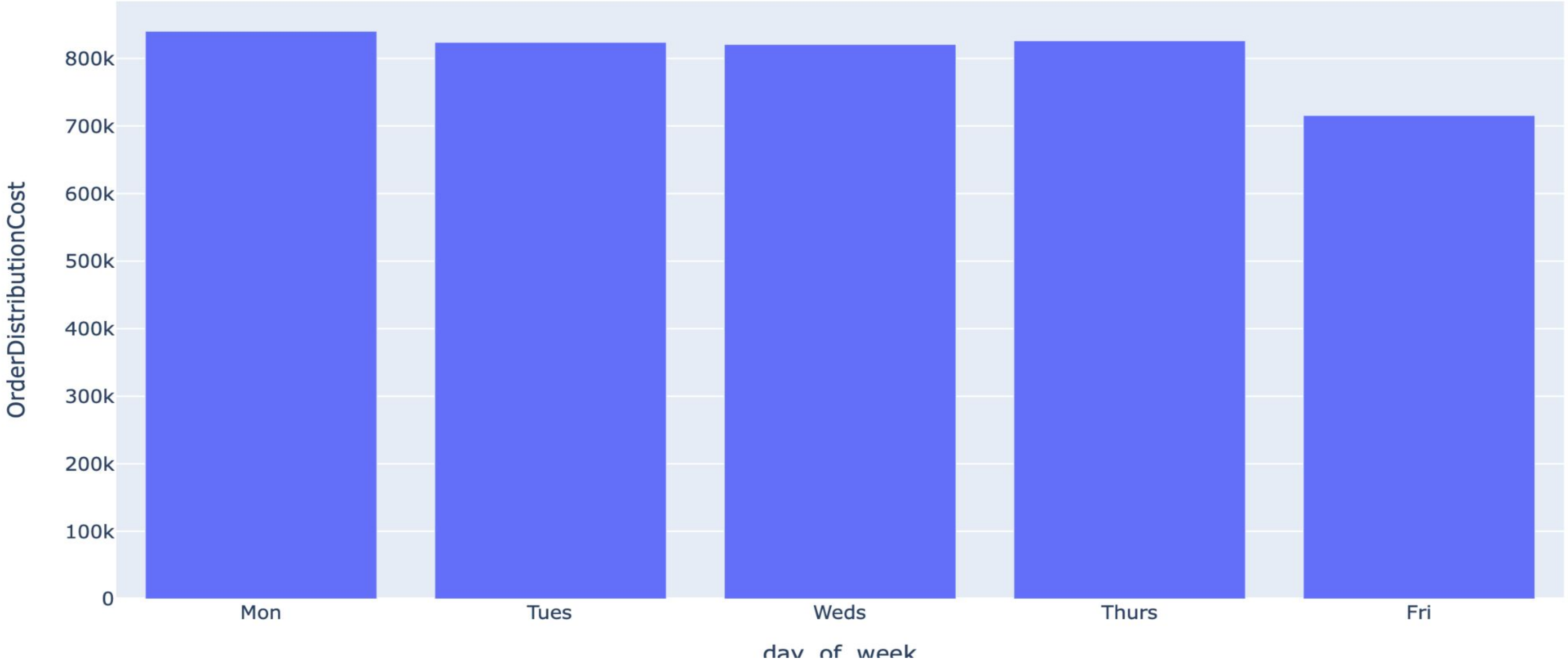
Order Total on Weekdays



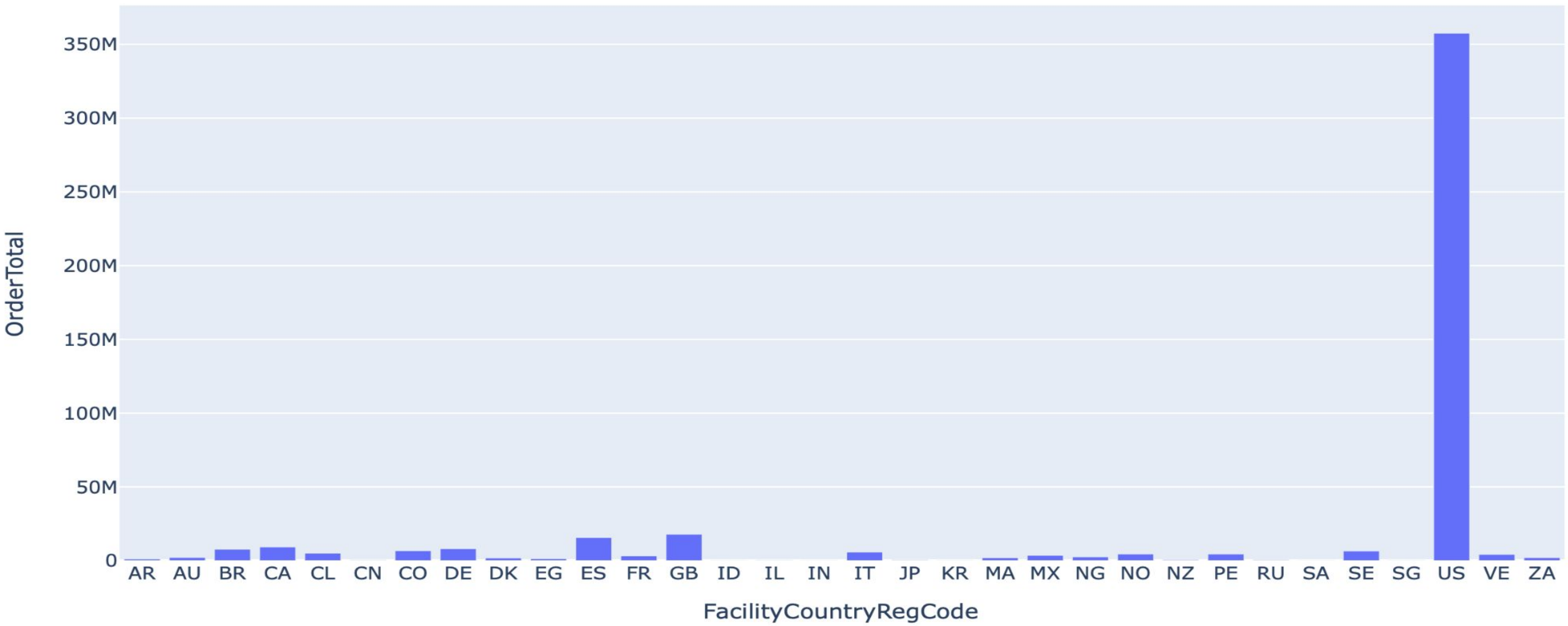
Production Cost on Weekdays



Order Distribution Cost on Weekdays



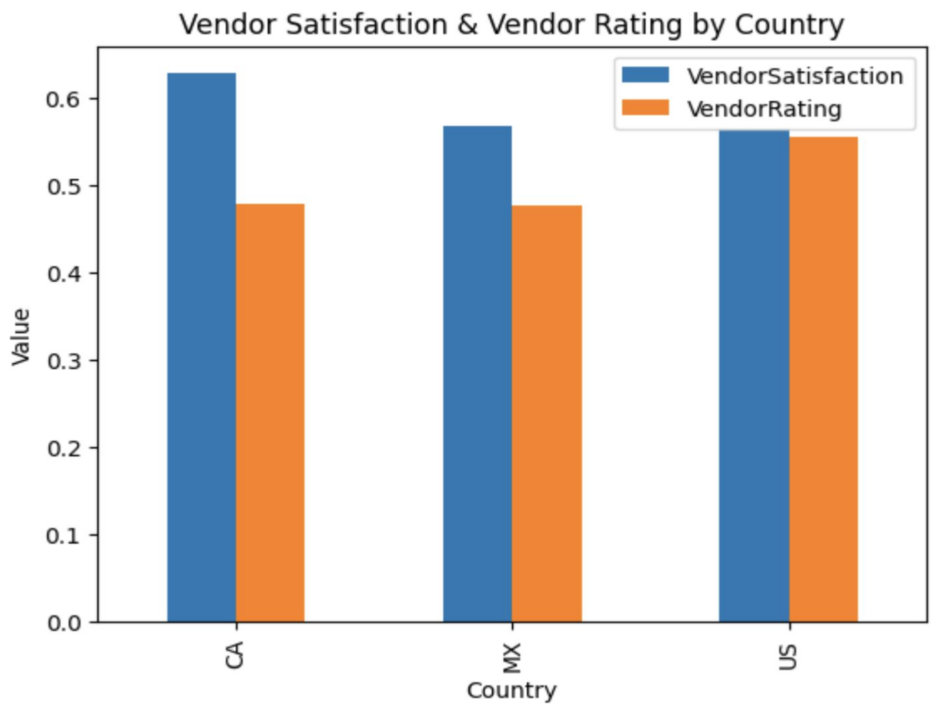
Order Total According to Facility Country

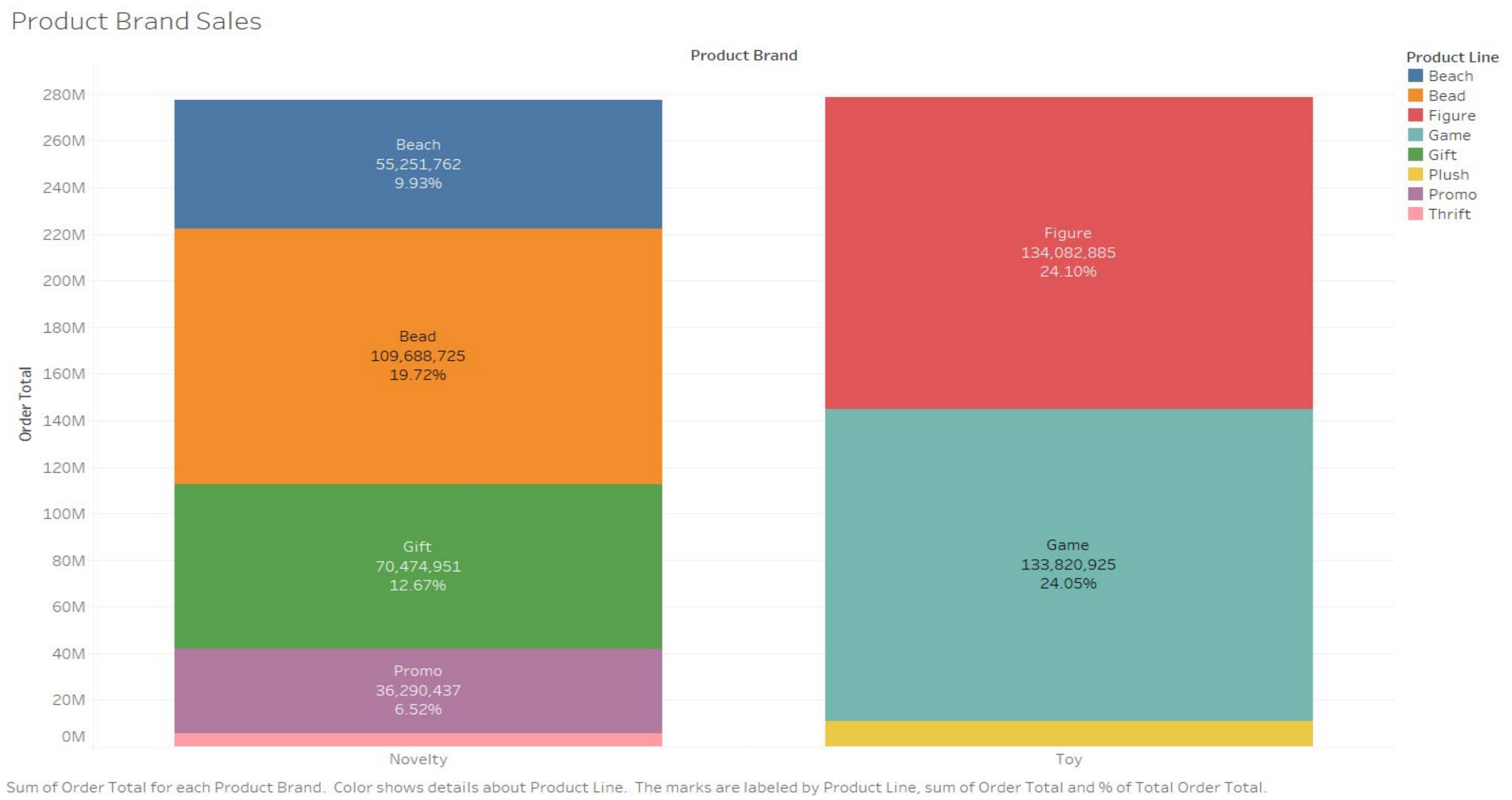


Vendor Satisfaction and Rating in NA countries

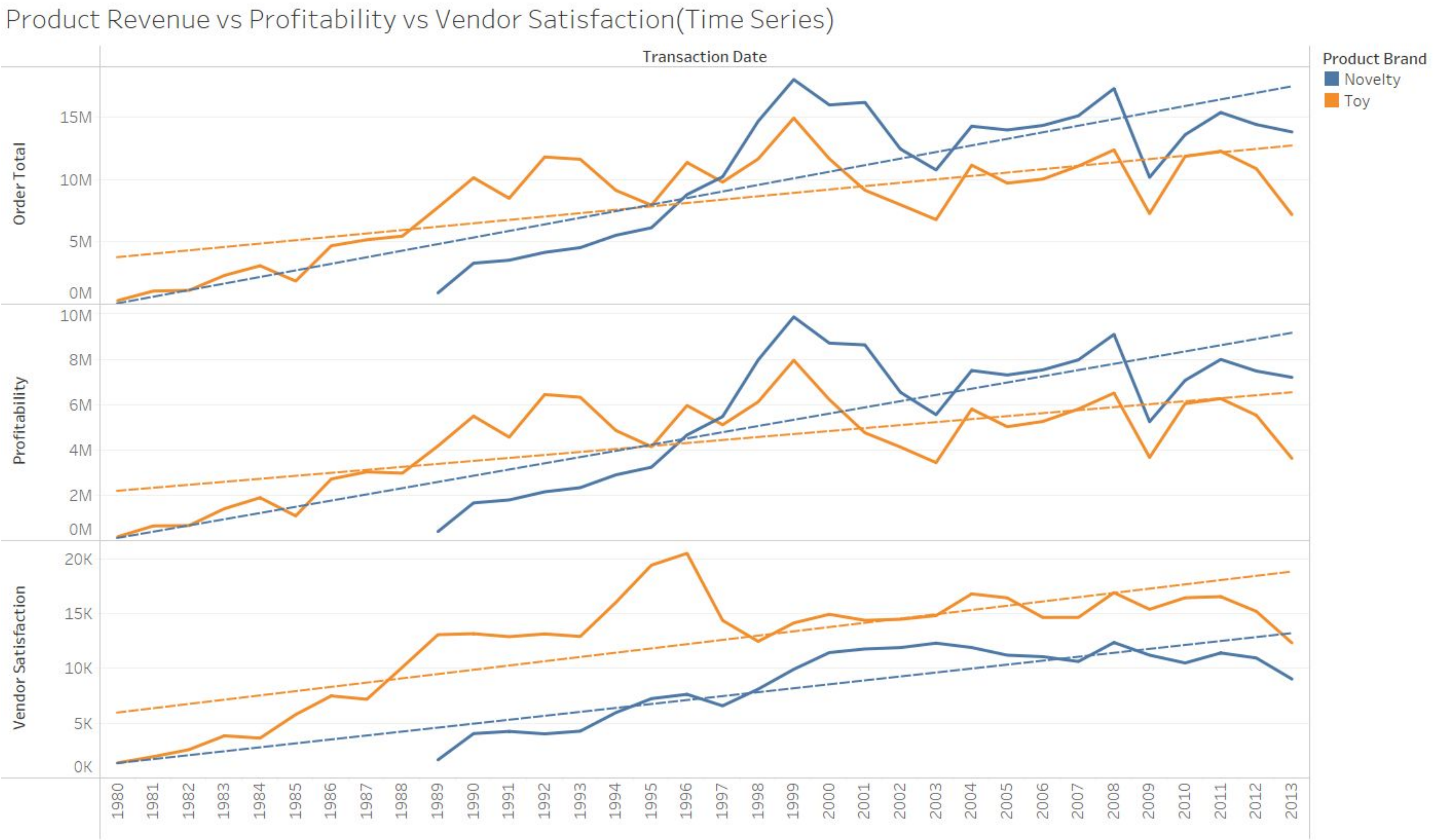
- This bar chart shows that the

vendor rating is high in Canada.



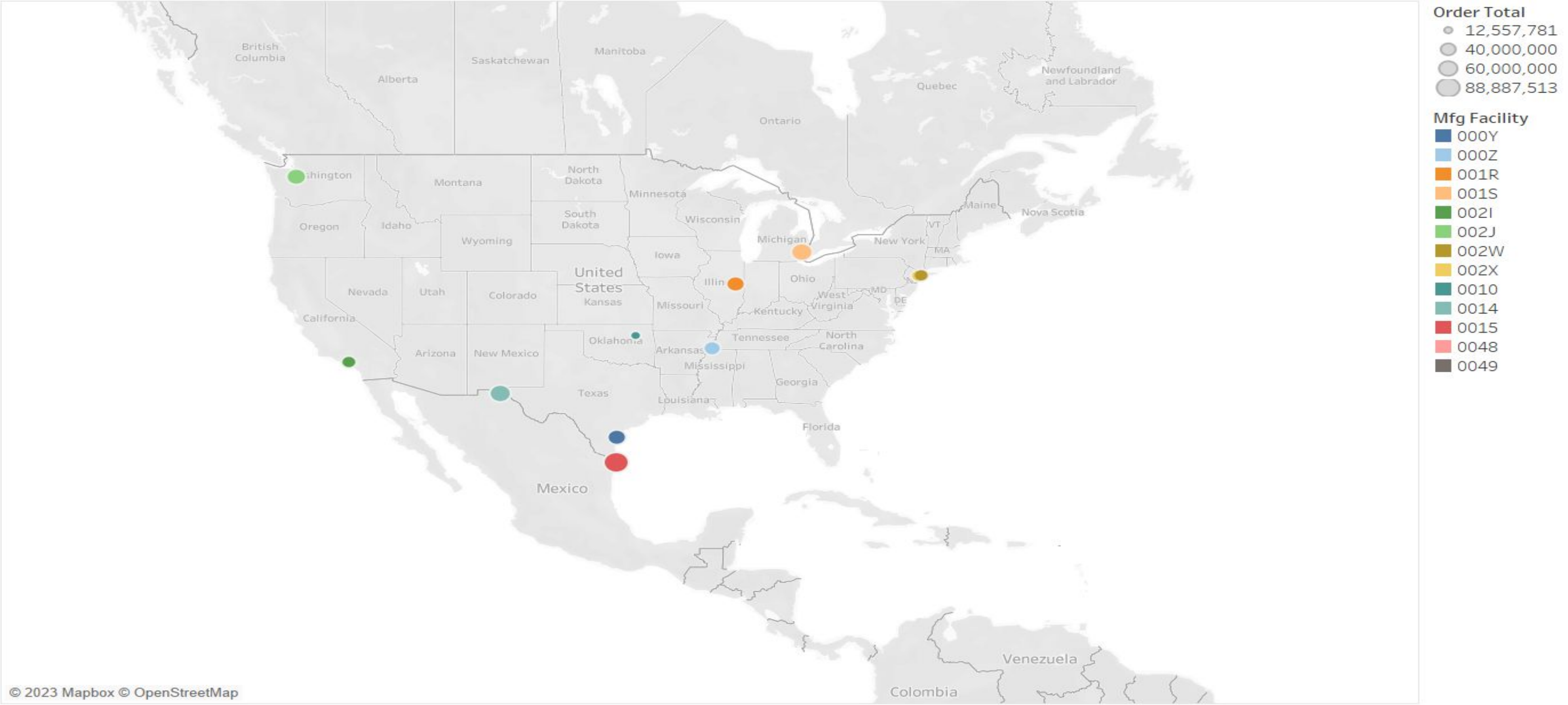


The main product type/brand are Toys and Novelty.



The trends of sum of Order Total, sum of Profitability and sum of Vendor Satisfaction for Transaction Date Year. Color shows details about Product Brand.

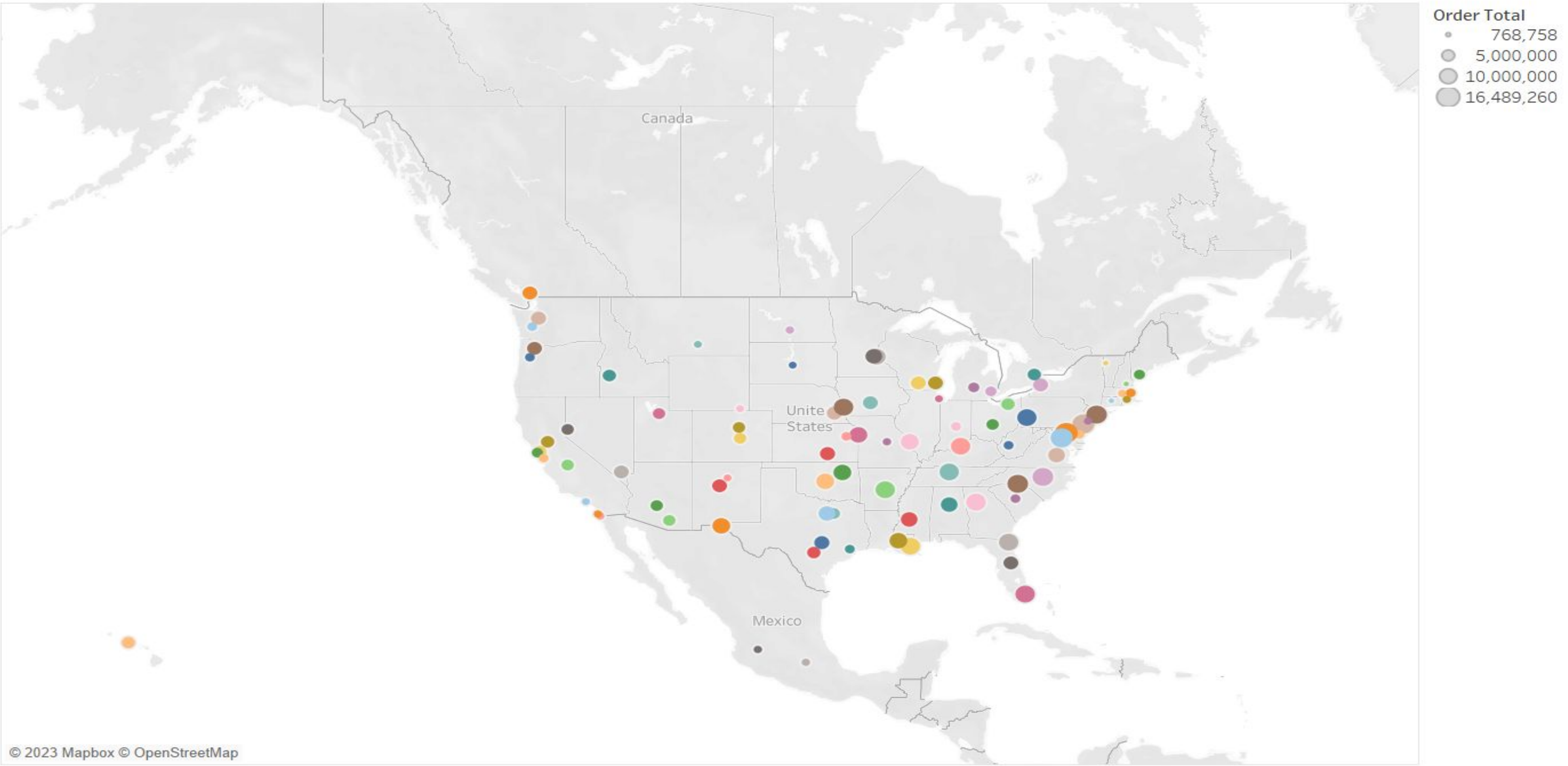
Manufacturing Facility vs Sales



Map based on average of Mfg Facility Lon and average of Mfg Facility Lat. Color shows details about Mfg Facility. Size shows sum of Order Total.

The company’s manufacturing facility in North America

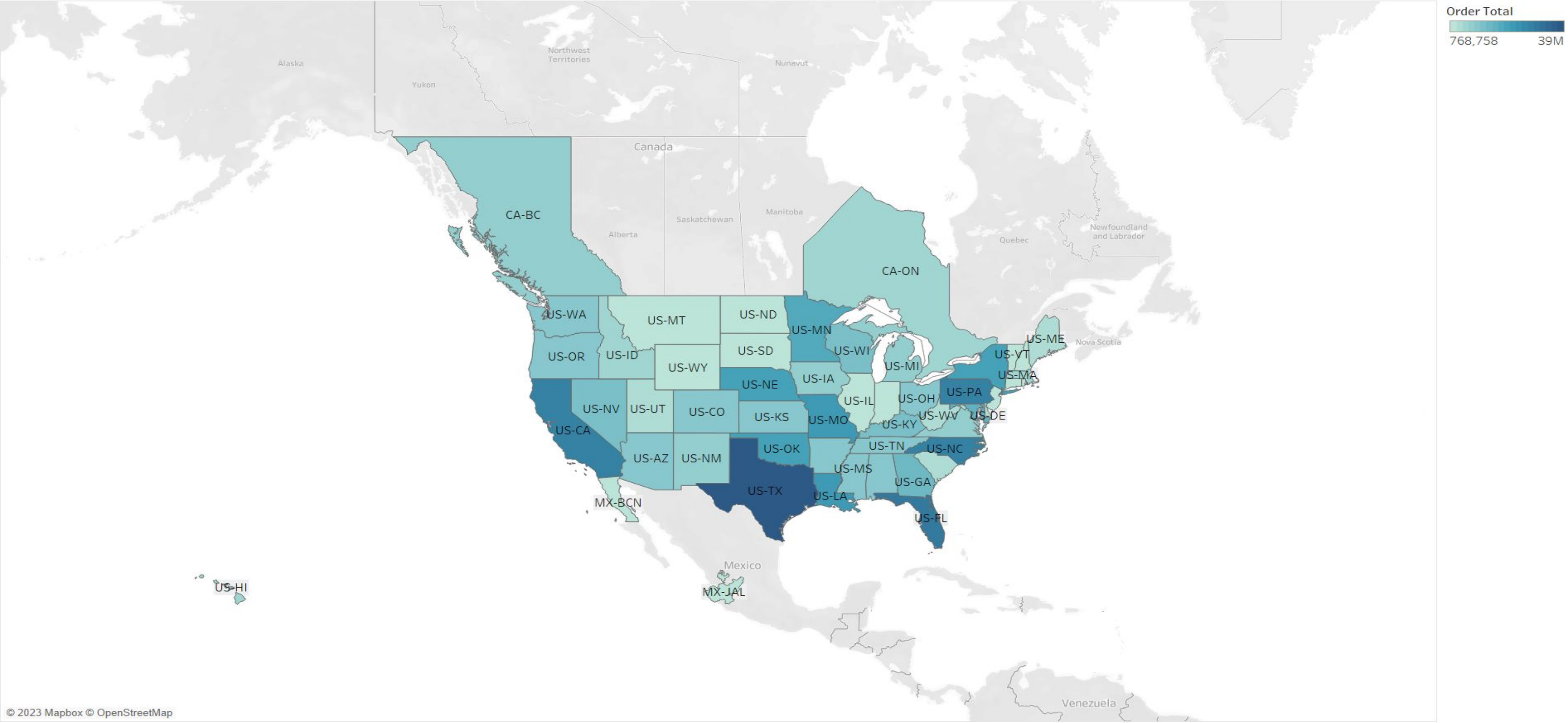
Facility Location vs Sales



Map based on average of Facility Lon and average of Facility Lat. Color shows details about Facility. Size shows sum of Order Total.

The company's Facility locations(NA)

Geographical Map - Sales by States



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Order Total. The marks are labeled by Facility State Prov. Details are shown for Facility Country Reg.

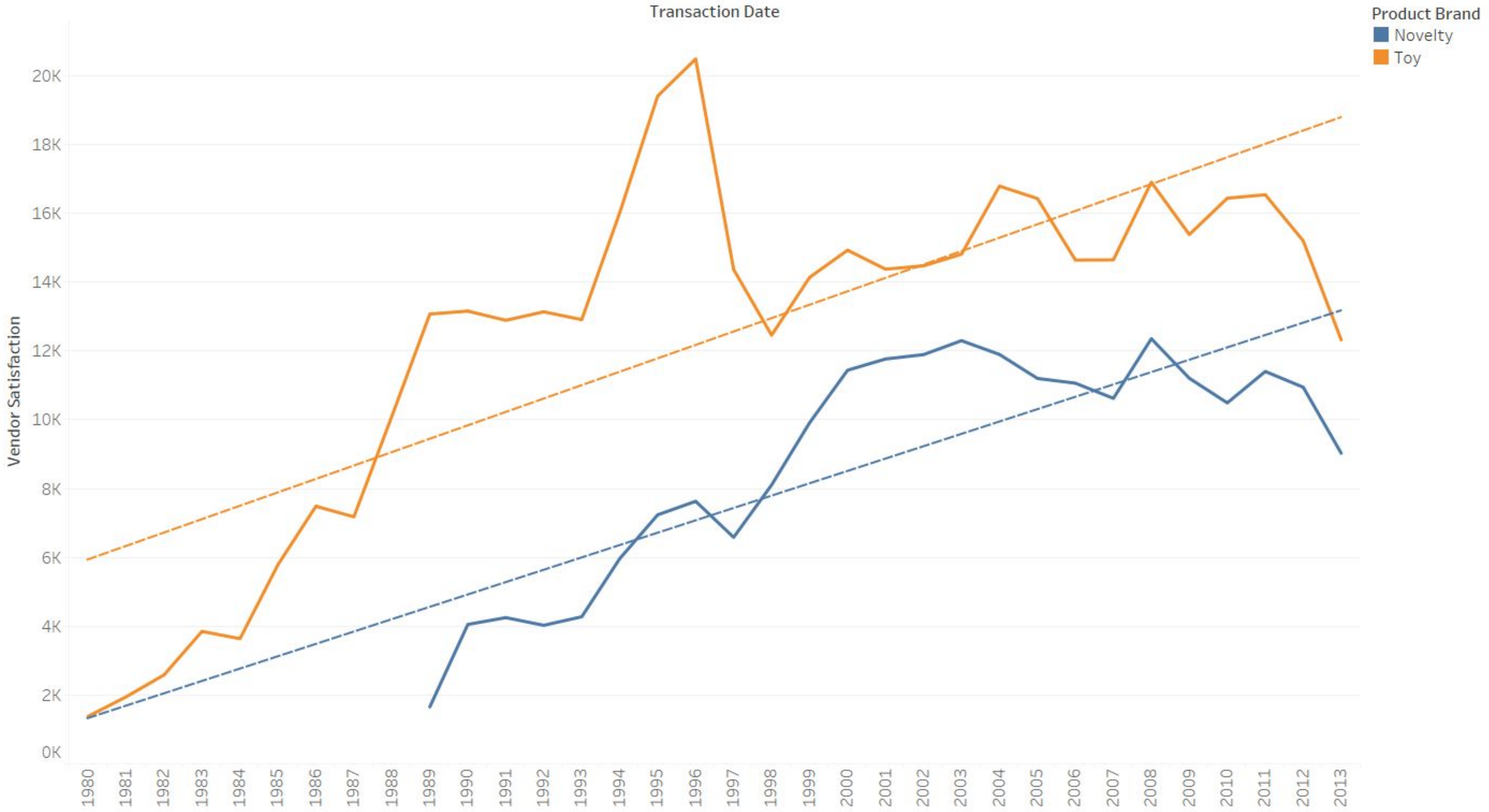
Company's Sales across States(NA)

Vendors vs Sales

Map based on average of Vendor Lon and average of Vendor Lat. Color shows details about Vendor. Size shows sum of Order Total.

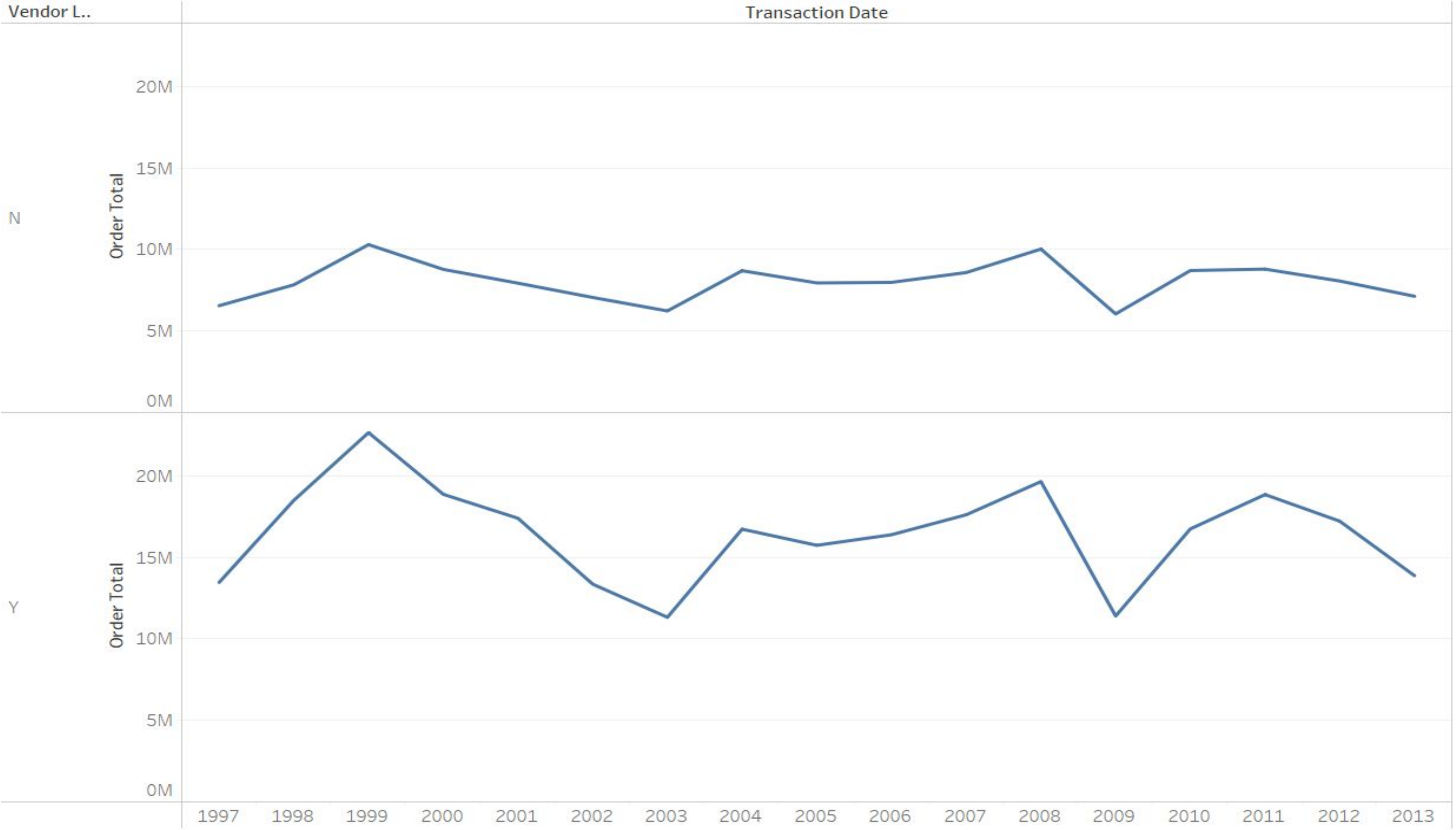
Company's Vendors Location (NA)

Vendor Satistfaction by Product Brand(Time Series)



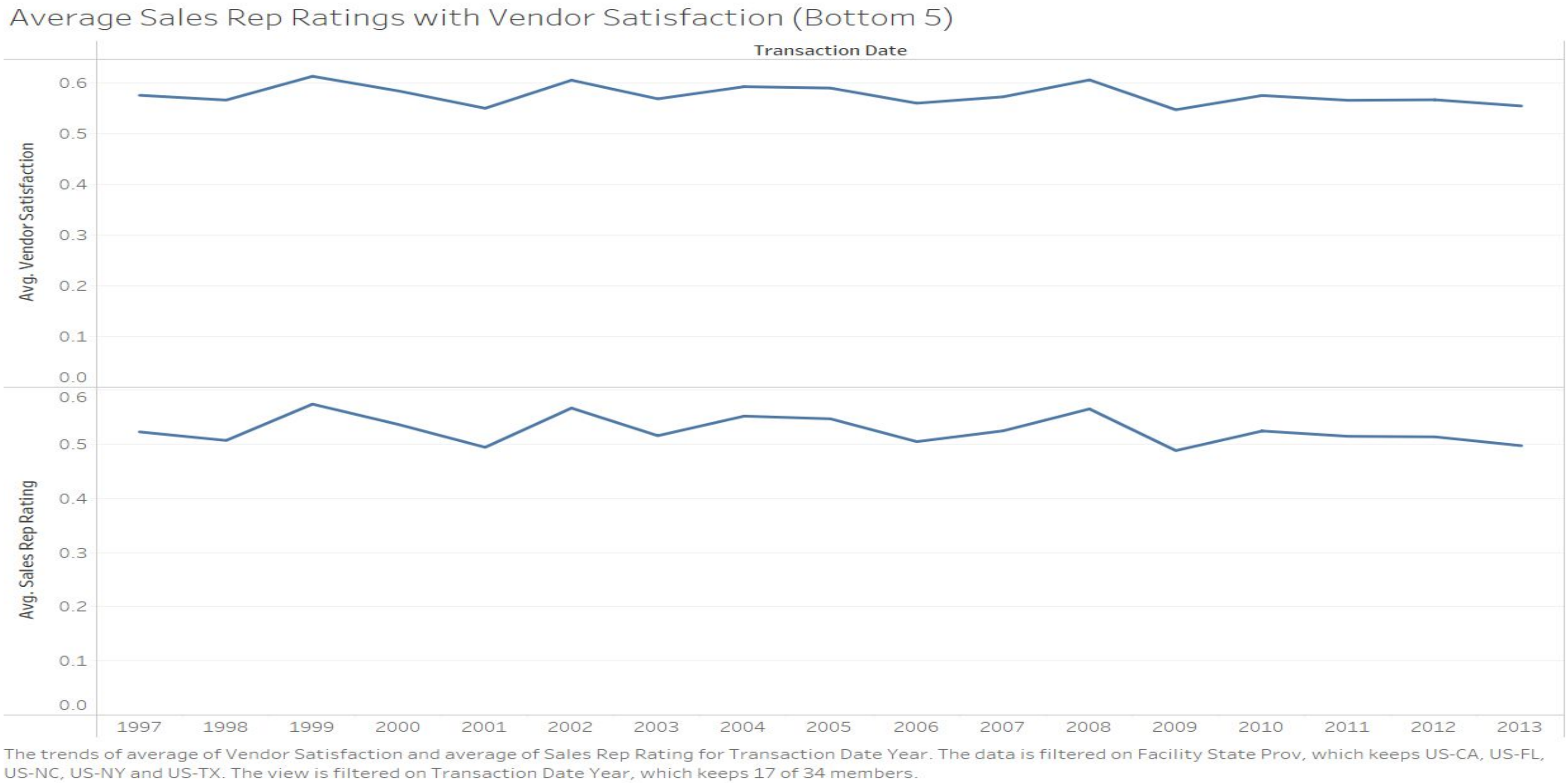
The trend of sum of Vendor Satisfaction for Transaction Date Year. Color shows details about Product Brand.

Vendor Loyalty Program vs Order Total



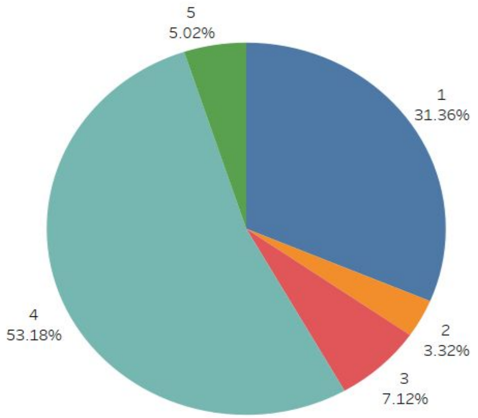
The trend of sum of Order Total for Transaction Date Year broken down by Vendor Loyalty Program. The data is filtered on Facility State Prov, which keeps 55 of 55 members. The view is filtered on Transaction Date Year, which keeps 17 of 34 members.

Average Sales Rep Rating with Vendor Satisfaction



Average Sales Rating and Average Vendor Satisfaction of the Bottom 5 companies.

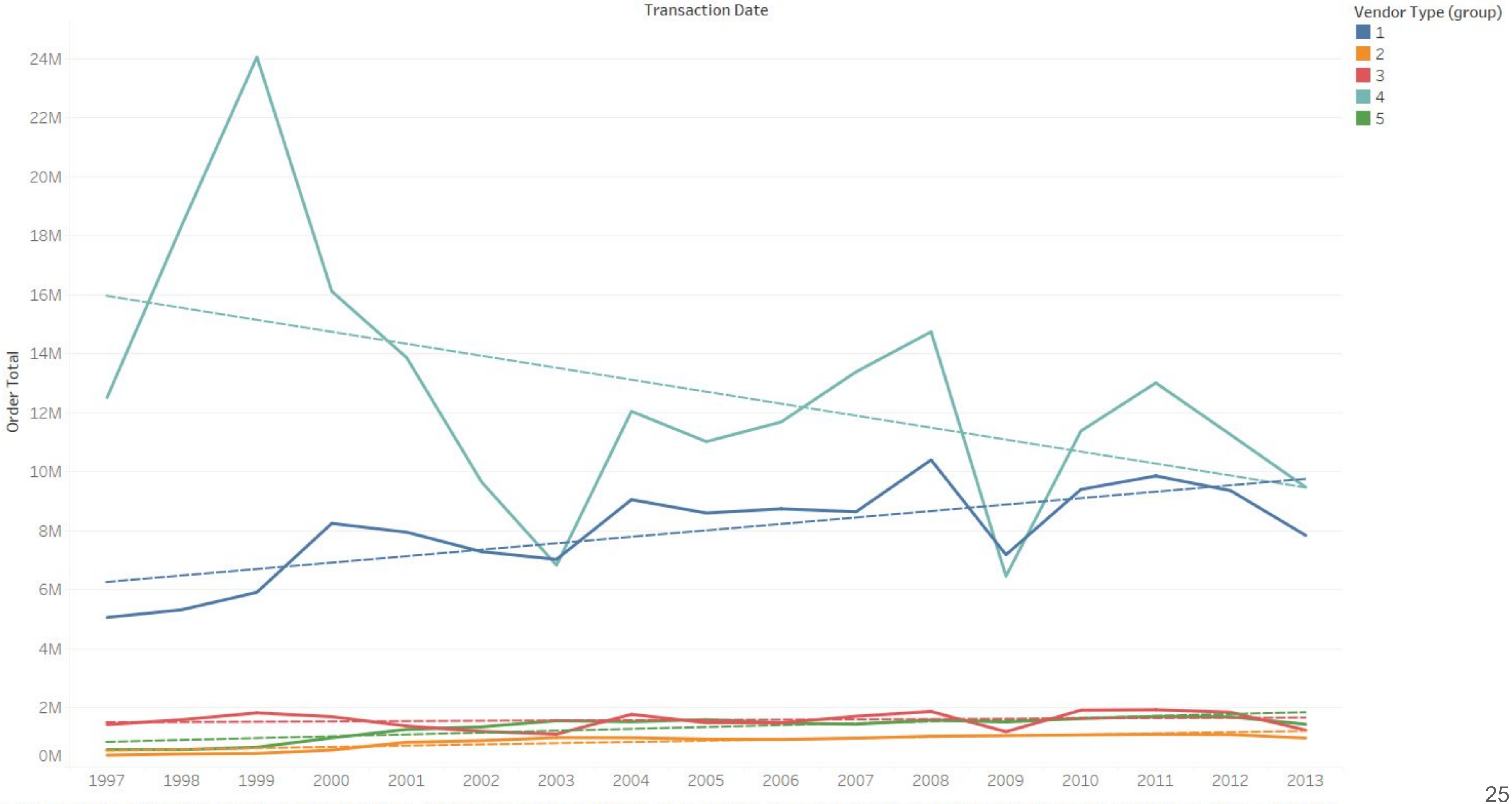
Vendor Type by Order Total



Vendor Type (group) and % of Total Order Total. Color shows details about Vendor Type (group). Size shows sum of Order Total. The marks are labeled by Vendor Type (group) and % of Total Order Total.

- There are five types of vendors: 1) Convenience Store, 2) Discount Store, 3) Department Store, 4) Kiosk & 5) Other.
- It can be observed that highest total orders are made through Kiosk.

Vendor Category Time Series



The trend of sum of Order Total for Transaction Date Year. Color shows details about Vendor Type (group). The data is filtered on Facility State Prov and Product Brand. The Facility State Prov filter keeps 55 of 55 members. The Product Brand filter keeps Novelty and Toy. The view is filtered on Transaction Date Year, which keeps 17 of 34 members.

Data Preparation

- We decided to go with analysis for the North America as it accounts for about 81% of the sales.
- We focused on 5 states that has decreasing order totals(revenue), trying to figure out the reason for the decrement. The 5 states are:
  - California
  - North Carolina
  - Florida
  - New York
  - Texas
- We also focused on 4 states that has increasing order totals(revenue), trying to figure out the reason for the increment. The 4 states are:
  - New Mexico
  - Nevada
  - Idaho
  - Maine
- We took data for the years starting from 1997 assuming after the internet boom(1996) people starting playing less with physical toys and that might be the cause of decreasing sales

Data Preparation(Contd..)

- Some columns from the dataset had incomplete rows, which was removed for the analysis. eg Vendor Ratings
- We had to change the data types for some columns i.e from binary to numeric, categorical etc for the analysis. eg Vendor Loyalty Program, Vendor Types
- Removed some columns for the analysis. e.g Order note, xyVendor Lon, xyVendor Lat etc
- Created few Subsets as per requirements for easier analysis. e.g: 5 states that has decreasing order totals, 4 states that has increasing order totals

Subsetting the data from 1997 based on the assumption

- TranscationDate was read by python as numeric data type
- Had to change the variable data type to datetime
- Then kept only the data where the date was after January 1997
- After subsetting the data was left with only 1048842 rows

```
[ ] df['TransactionDate'] = pd.to_datetime(df['TransactionDate'])
```

```
[ ] df['TransactionDate']

0          1999-01-04
1          1989-01-02
2          1989-01-02
3          1989-01-02
4          2000-01-03
              ...
1416053    2013-10-30
1416054    2013-10-30
1416055    2013-10-30
1416056    2013-10-30
1416057    2013-10-30
Name: TransactionDate, Length: 1416058, dtype: datetime64[ns]
```

```
[ ] df = df[df['TransactionDate'] >= '1997-01-01']
```

```
df['TransactionDate']

0          1999-01-04
4          2000-01-03
5          2000-01-03
6          2000-01-03
7          2000-01-03
              ...
1416053    2013-10-30
1416054    2013-10-30
1416055    2013-10-30
1416056    2013-10-30
1416057    2013-10-30
Name: TransactionDate, Length: 1048842, dtype: datetime64[ns]
```

```
df['TransactionDate'].describe()

count                 1048842
unique                   4391
top       2011-06-07 00:00:00
freq                      479
first     1997-01-01 00:00:00
last      2013-10-30 00:00:00
Name: TransactionDate, dtype: object
```

```
[ ] df['FacilityContinent'].replace(np.nan, 'North_America', inplace=True)
```

```
df['FacilityContinent']
```

```
0          North_America
4                     EU
5                     EU
6                     EU
7                     EU
                ...
1416053    North_America
1416054    North_America
1416055    North_America
1416056    North_America
1416057    North_America
Name: FacilityContinent, Length: 1048842, dtype: object
```

Subsetting the data only for North America

- Had to rename NA as North_America as the python read NA as null value
- Then removed data for other continent and left with only North America Data
- After subsetting the data has 773858 rows

```
[ ] necesssary_values = ['North_America']
    df = df[df['FacilityContinent'].isin(necesssary_values)]
```

```
[ ] df['FacilityContinent']
```

```
0          North_America
21         North_America
22         North_America
23         North_America
24         North_America
                ...
1416053    North_America
1416054    North_America
1416055    North_America
1416056    North_America
1416057    North_America
Name: FacilityContinent, Length: 773858, dtype: object
```

Dropping the columns that are not relevant to the Analysis and Model

```
[ ] columns_to_drop = ['OrderNote', 'FacilityLat','FacilityLon','FacilityDateOpened', 'FacilityDateClosed', 'FacilityContinentLat',
                       'FacilityContinentLon','FacilityCountryRegLat', 'FacilityCountryRegLon',
                       'FacilityStateProvLat', 'FacilityStateProvLon','FacilityCityLat', 'FacilityCityLon','VendorLat', 'VendorLon',
                       'MfgFacilityLat', 'MfgFacilityLon',]
    df = df.drop(columns_to_drop, axis=1)
```

```
df
```

	Order	OrderTotal	OrderProductCost	OrderDistributionCost	OrderSalesCost	OrderMarketingCost	MfgBatch	MfgBatchSKU	SalesRep	Facility	...
0	00000001	333.113707	111.037902	1.897610	2.314892	10.873238	00000003	017308	023S	002M	...
21	0000000A	419.906722	167.962689	2.420791	4.296466	17.458488	0000000V	01730B	01QK	0025	...
22	0000000F	129.935263	51.974105	2.538320	4.807432	4.881309	0000000X	017B3Z	005U	000S	...
23	0000000G	346.494036	173.247018	4.912435	1.271986	18.658396	0000000Y	0172XF	00SC	001I	...
24	0000000P	85.507604	42.753802	3.607503	1.271986	3.001294	00000010	017J1W	00SC	001I	...
...	...	...	...	...	...	...	...	...	...	...	...
1416053	000120MI	1601.983076	640.793230	8.116594	6.777258	69.164760	0001GQ8U	00LV6X	00JJ	001E	...
1416054	000120MG	1601.983076	800.991538	14.939014	6.528912	87.578755	0001GQ8V	00LV9O	00V8	001J	...
1416055	000120MF	1586.563727	634.625491	8.790131	7.131529	69.370656	0001GQ8W	00M32B	00K6	001C	...
1416056	000120ML	1201.487307	480.594923	6.833819	6.904941	52.481031	0001GQ90	00LV4A	00JJ	001E	...
1416057	000120MK	1269.031601	634.515800	8.249153	6.862072	68.942402	0001GQ91	00M37T	00JR	001F	...

773858 rows × 40 columns

After dropping, the data set has 773858 rows and 40 columns

Subsetting the data for states with decreasing revenue for regression analysis model (Model-D)

```
[ ] Df_decending = df_main
```

```
[ ] condition = Df_decending['FacilityStateProv'].isin(['US-CA', 'US-FL','US-NC','US-NY','US-TX'])
    Df_decending = Df_decending[condition]
```

```
Df_decending
```

	Order	OrderTotal	OrderProductCost	OrderDistributionCost	OrderSalesCost	OrderMarketingCost	MfgBatch	MfgBatchSKU	SalesRep	Facility	...
22	0000000F	129.935263	51.974105	2.538320	4.807432	4.881309	0000000X	017B3Z	005U	000S	...
26	0000000R	77.961158	25.987053	1.794254	1.363753	1.685789	00000014	017B6P	023Q	002M	...
28	0000000Q	108.279386	43.311754	0.876603	4.941638	3.108800	00000017	017J1X	01XB	001W	...
29	00000018	541.396931	216.558772	3.248573	5.197438	22.496476	00000018	00M2WR	01XB	001W	...
132	0000004J	409.471255	163.788502	5.708276	6.336897	17.262075	0000005A	0172XC	01A7	000O	...
...	...	...	...	...	...	...	...	...	...	...	...
1409462	00011SS9	661.840403	220.613468	2.927153	6.515758	24.139250	0001GG38	00M329	0116	0008	...
1409464	00011SSA	859.771862	286.590621	4.611474	4.698227	30.298938	0001GG3A	00M2WS	018G	000A	...
1409466	00011SSD	287.975413	95.991804	3.241796	7.140102	9.551817	0001GG3C	00LV1L	0116	0008	...
1409467	00011SSE	1408.570494	469.523498	6.393707	6.913937	50.332456	0001GG3D	00LVCP	011E	0007	...
1409468	00011SSG	401.022366	160.408947	3.182272	5.981692	16.629550	0001GG3F	00LV75	008Q	000S	...

175994 rows × 40 columns

This subset data set has 175994 rows and 40 columns for regression analysis

Subsetting the data for states with increasing revenue for regression analysis model (Model-I)

```
[ ] condition = Df_acending['FacilityStateProv'].isin(['US-PA','US-NV','US-ID','US-ME'])
    Df_acending = Df_acending[condition]
```

```
Df_acending
```

	Order	OrderTotal	OrderProductCost	OrderDistributionCost	OrderSalesCost	OrderMarketingCost	MfgBatch	MfgBatchSKU	SalesRep	Facility	...
995	00000107	683.869808	227.956603	4.376926	1.588944	23.990509	00000182	017APX	024C	002N	...
1230	0000018O	344.885515	114.961838	6.153978	8.001744	11.880321	000001IC	01738J	024I	002O	...
1231	00000190	364.730564	121.576855	2.777262	4.624597	12.719263	000001ID	0172XD	0246	002N	...
1233	0000018W	91.182641	30.394214	5.308302	8.001744	1.368394	000001IF	017B3X	024I	002O	...
1234	00000199	433.450796	144.483599	2.795831	4.540398	14.647955	000001IG	017JIM	0246	002N	...
...	...	...	...	...	...	...	...	...	...	...	...
1399352	00011IBV	2868.912667	956.304222	13.758486	7.553802	103.276418	0001G2AY	00LV9W	0282	002O	...
1399393	00011IDI	4562.306447	1520.768816	18.258918	7.096117	166.326740	0001G2D8	00M37S	0282	002O	...
1399394	00011IDG	165.943938	73.752861	1.369304	6.056749	6.385039	0001G2DA	00MAMP	01FH	002S	...
1399396	00011IDJ	3113.570650	1383.809178	14.342269	6.005710	150.839958	0001G2DD	00LV6Y	01FH	002S	...
1399398	00011IDO	240.297461	80.099154	2.612832	5.966548	6.908327	0001G2DK	00MASB	026Z	002N	...

59780 rows × 40 columns

This subset data set has 59780 rows and 40 columns for regression analysis

We have Model-D and Model-I so that we can compare the models - collect findings and give suggestion to states in Model-D to increase revenue in the future

Data Understanding

- The dataset contains 57 variables as columns and over 1,416,058 observations as rows.
- The data can be categories into Manufacturing, Facility location, Costs involved in distribution/marketing/sales, Market Penetration, Vendor types, Product types, Vendor information, Sales Rep information and Revenue
- The data types consist of Characters, Date and Numeric.
- The data contains missing values in certain fields that needs to be analyzed and removed.
- Used different descriptive statistics to understand the mean, standard deviation, different quartiles.

```
In [12]: rows=toyUS.shape[0]
```

```
In [13]: column=toyUS.shape[1]
```

```
In [15]: print ('Number of rows and columns', rows, '&', column)
```

```
Number of rows and columns 1141074 & 58
```

Checking Rows and Columns

```
toyUS.dtypes
```

```
Unnamed: 0                 int64
Order                     object
OrderTotal               float64
OrderProductCost         float64
OrderDistributionCost    float64
OrderSalesCost           float64
OrderMarketingCost       float64
OrderNote                 object
MfgBatch                  object
MfgBatchSKU               object
SalesRep                  object
Facility                  object
FacilityLat              float64
FacilityLon              float64
FacilityDateOpened        object
```

Checking DataTypes for different Columns

```
null_values = Df_toy['VendorSatisfaction'].isnull().sum()
```

```
print(null_values)
```

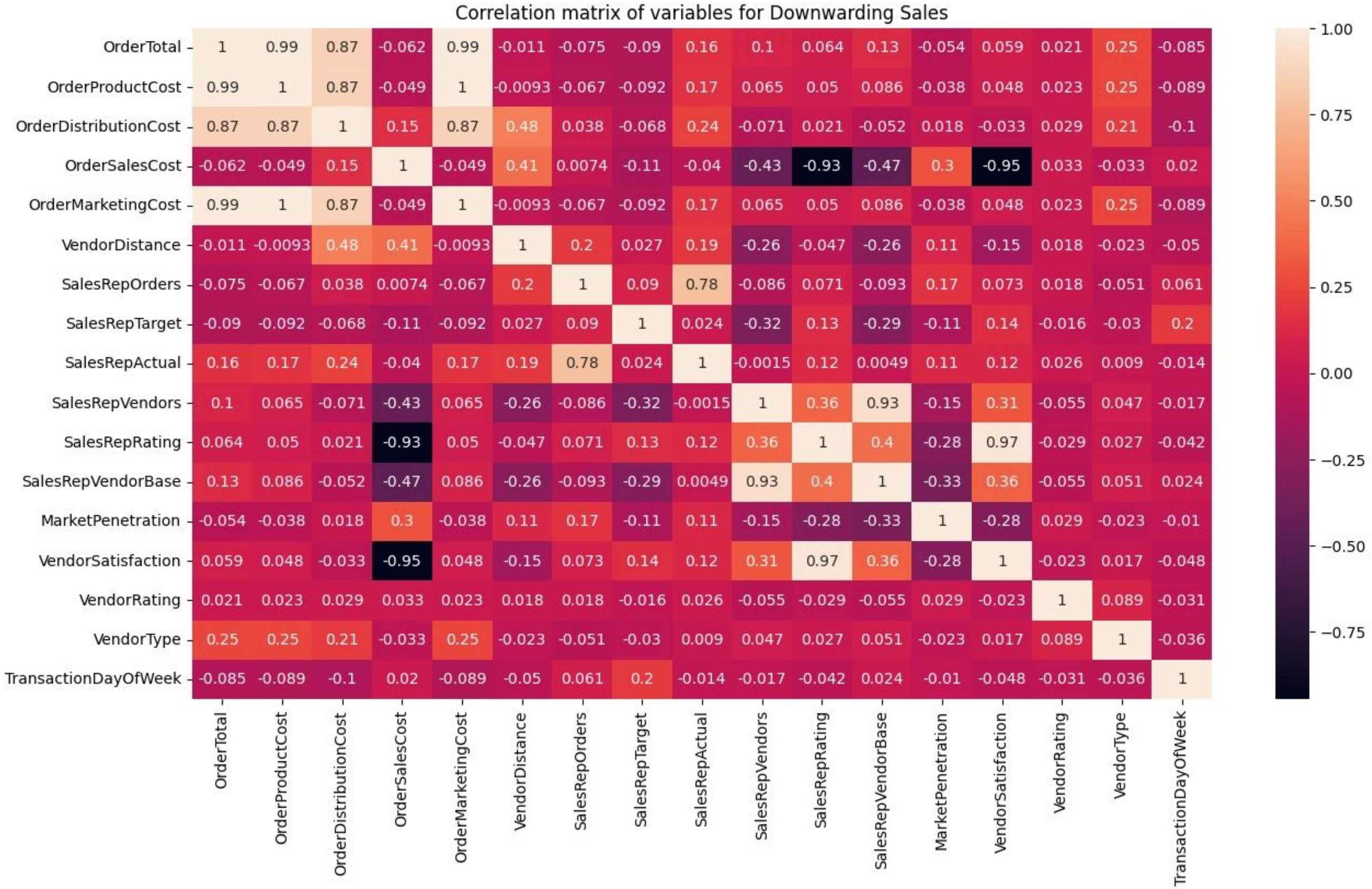
```
218
```

Checking Null Values in the Column

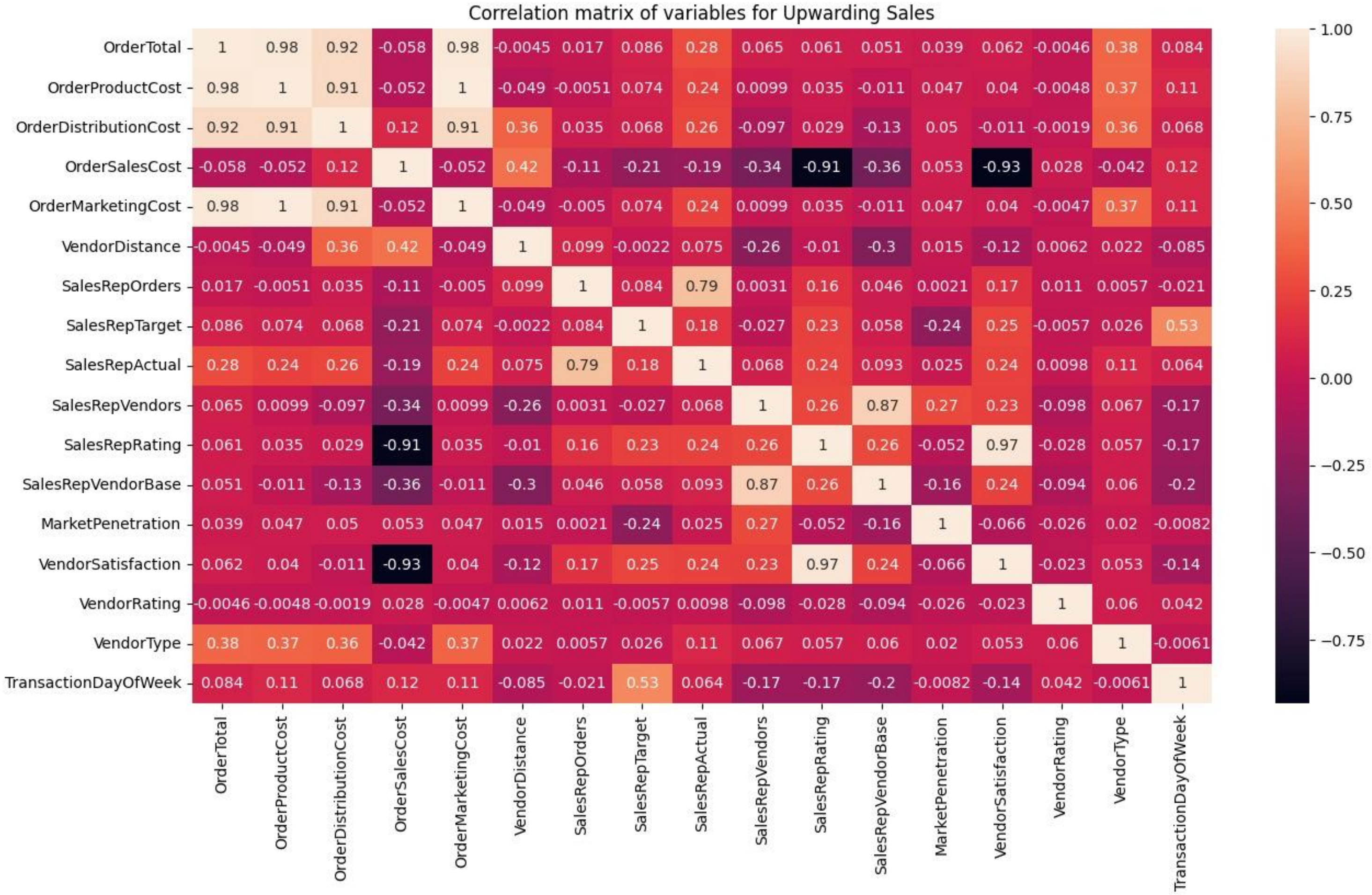
```
print(toyUS.describe())
```

```
       Unnamed: 0  OrderTotal  OrderProductCost  OrderDistributionCost  \
count 1141074.000 1141074.000       1141074.000            1141074.000
mean   745390.651     487.557           198.020                  4.995
std    407431.772     689.400           277.332                  3.511
min         0.000       1.825             0.913                  0.041
25%    413298.250      71.699            29.710                  2.521
50%    763890.500     224.697            91.183                  4.094
75%   1092931.750     628.542           255.688                  6.500
max   1416057.000   16903.930          5695.940                 57.597
```

Checking Descriptive Statistics for different Columns



Correlation between the variables from decreasing orders total states

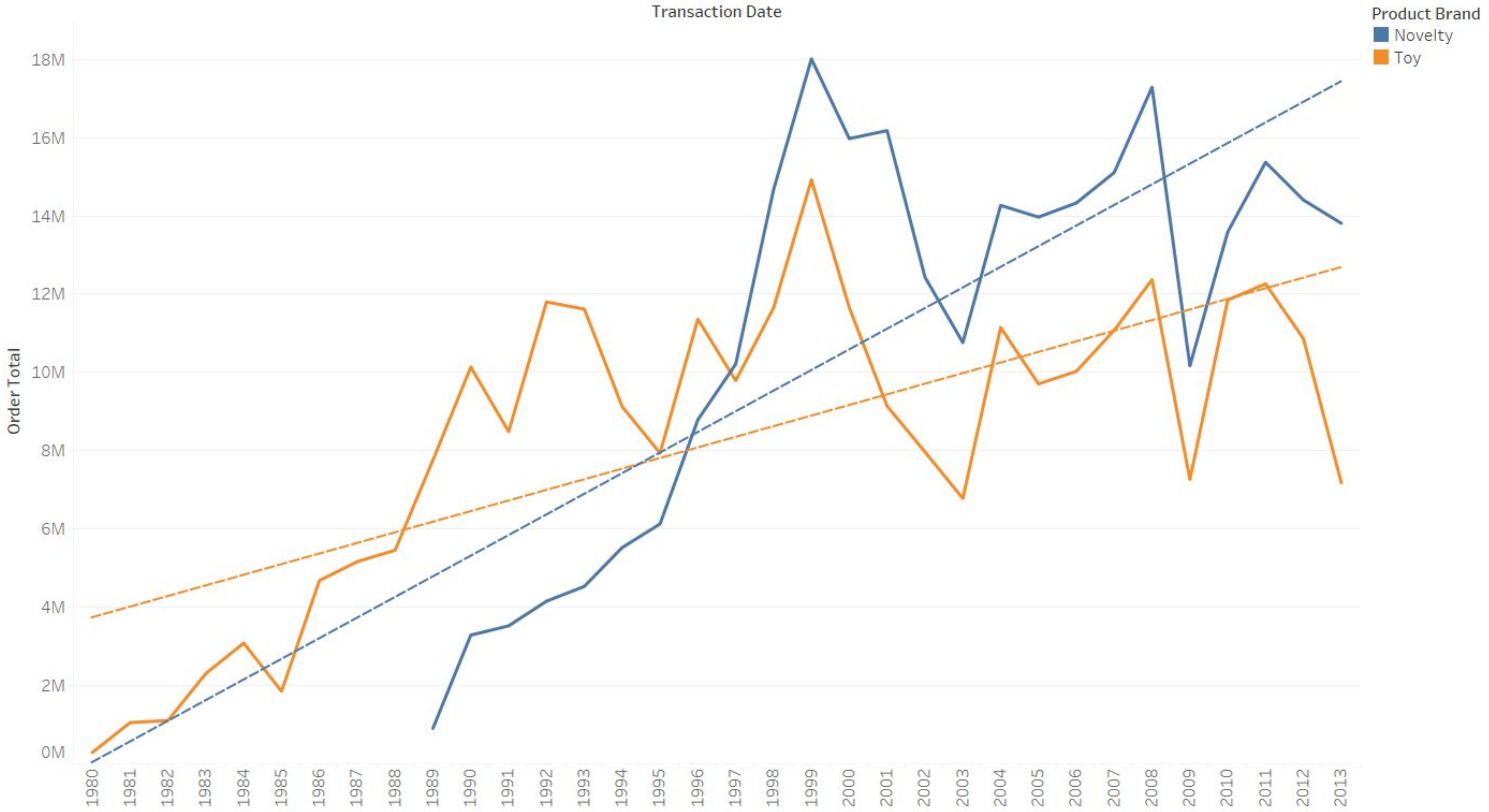


Correlation between the variables from increasing orders total states

Correlation between the variables (Vendor Satisfaction)

Product Understanding: Revenue

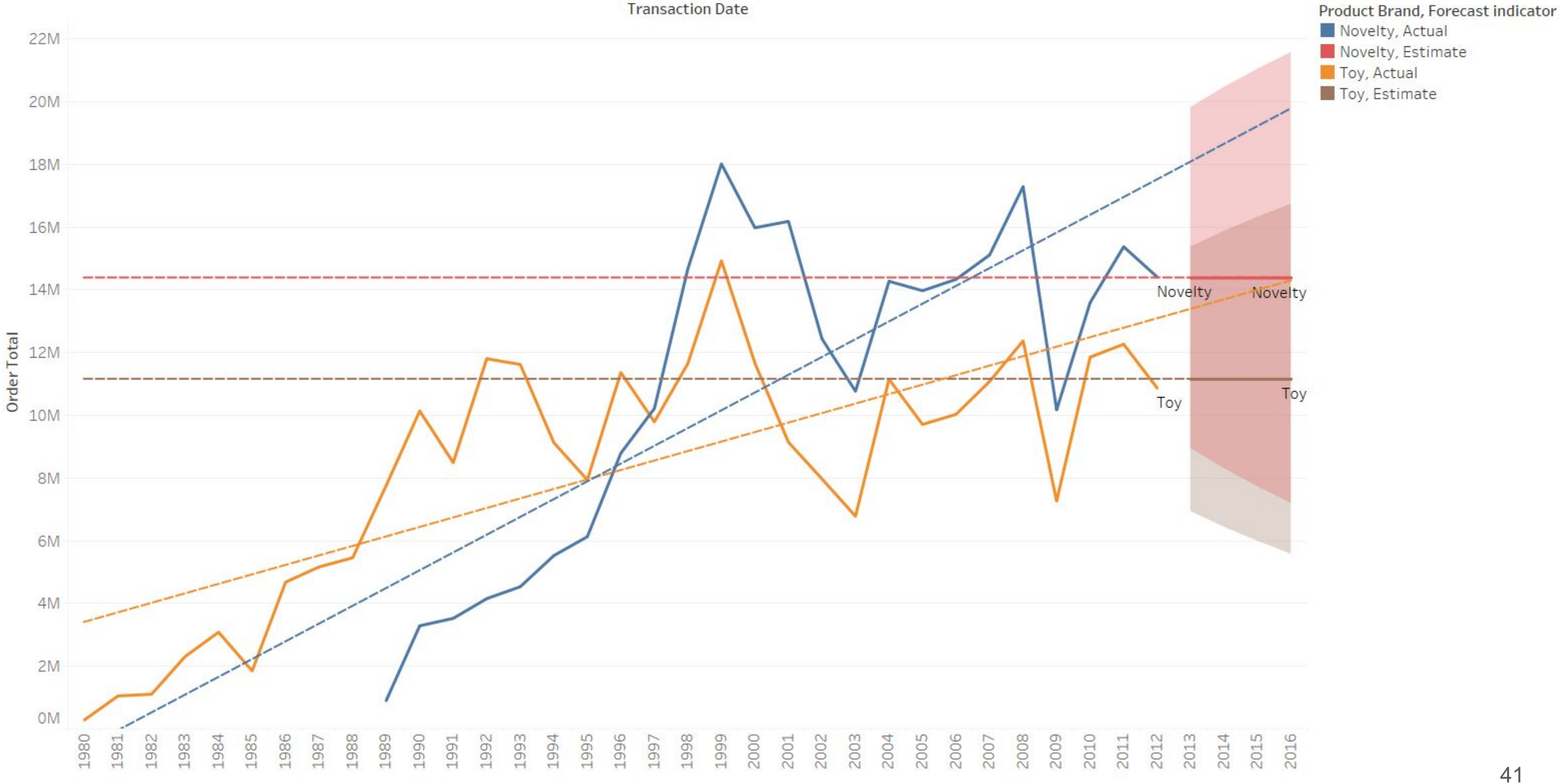
Product Revenue(Time Series)



The trend of sum of Order Total for Transaction Date Year. Color shows details about Product Brand.

0

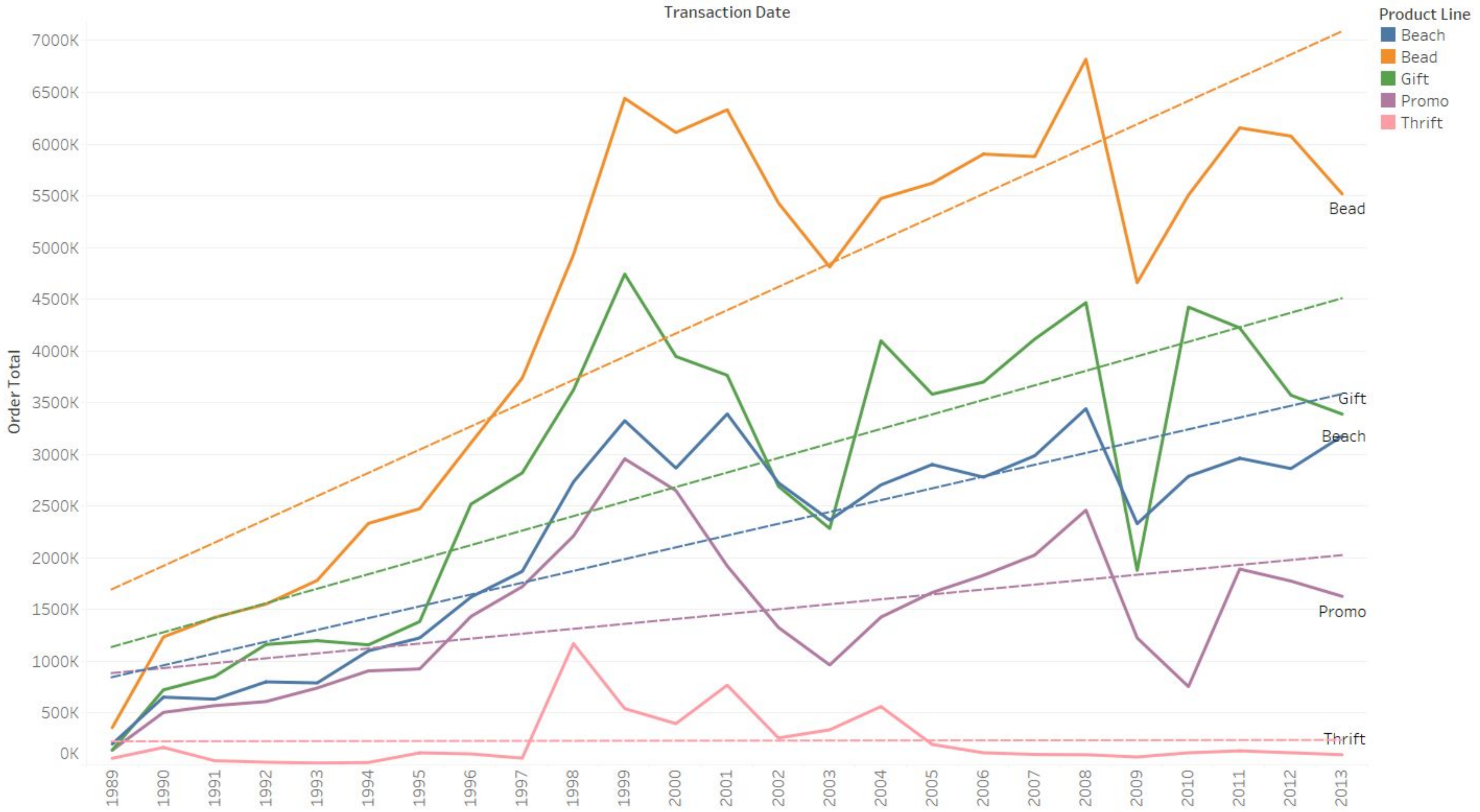
Product Brand (Time Series) Actual vs Estimate



The trend of sum of Order Total (actual & forecast) for Transaction Date Year. Color shows details about Product Brand and Forecast indicator. The marks are labeled by Product Brand.

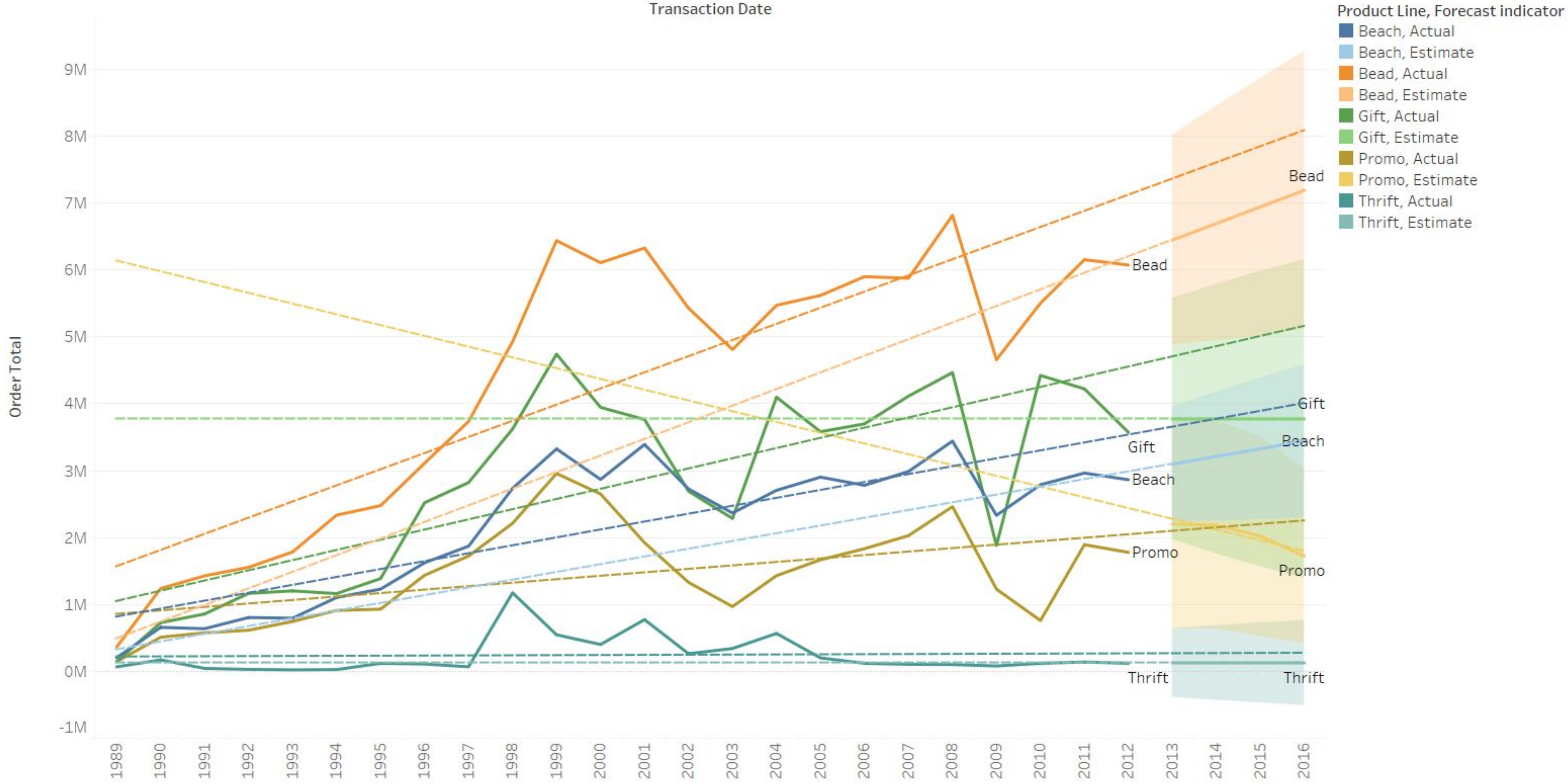
Product Brand	Actual	Estimate
Novelty	Order Total = 527172*Year of Transaction Date + -1.04377e+09 R-Squared: 0.58825 P-value: < 0.0001	Order Total = 0*Year of Transaction Date + 1.43749e+07 R-Squared: 1 P-value: N/A
Toy	Order Total = 302341*Year of Transaction Date + -5.95238e+08 R-Squared: 0.562663 P-value: < 0.0001	Order Total = 0*Year of Transaction Date + 1.11476e+07 R-Squared: 1 P-value: N/A

Novelty - Product Brand (Time Series)



The trend of sum of Order Total for Transaction Date Year. Color shows details about Product Line. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Novelty.

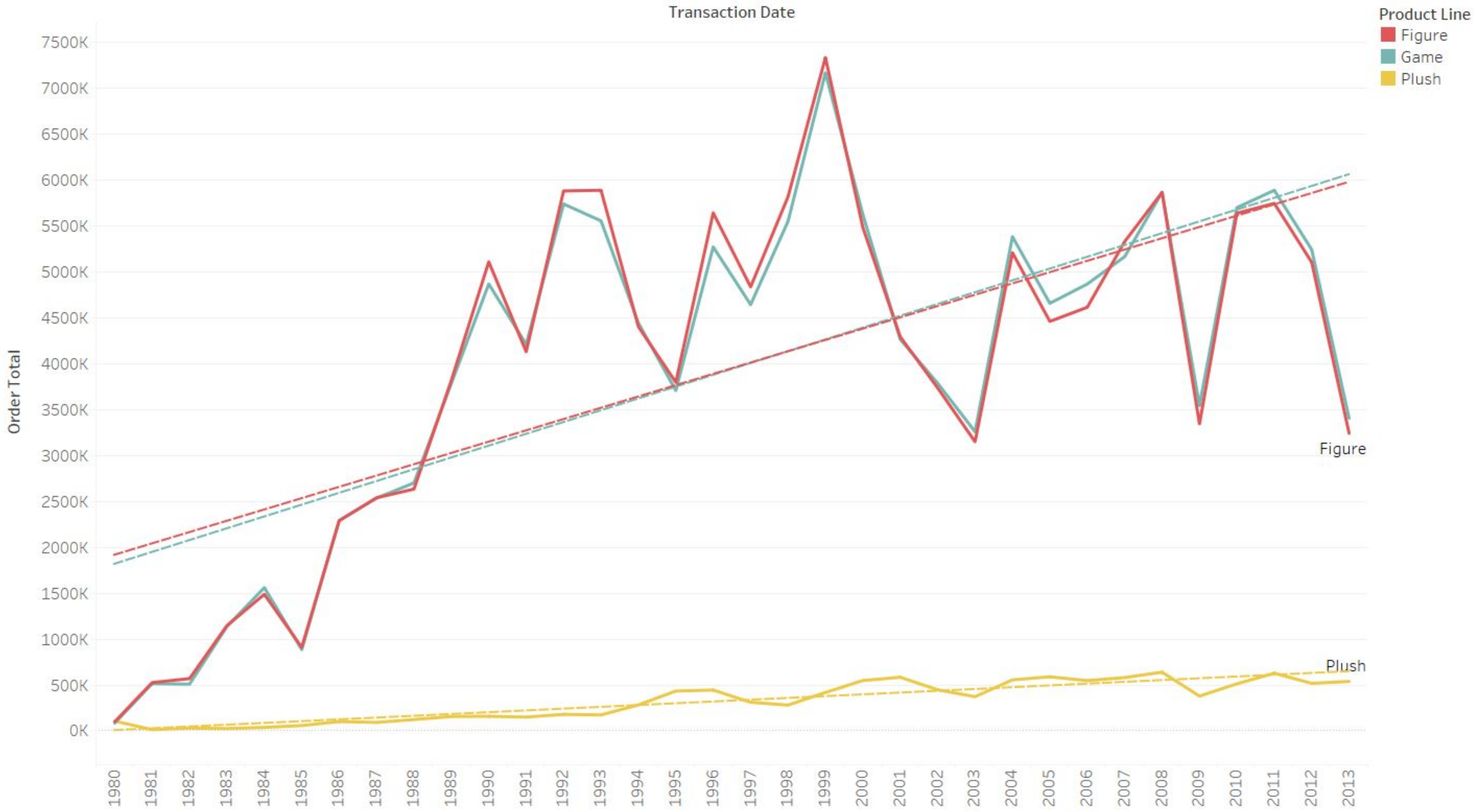
Novelty - Product Brand (Time Series) Actual vs Estimation



The trend of sum of Order Total (actual & forecast) for Transaction Date Year. Color shows details about Product Line and Forecast indicator. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Novelty.

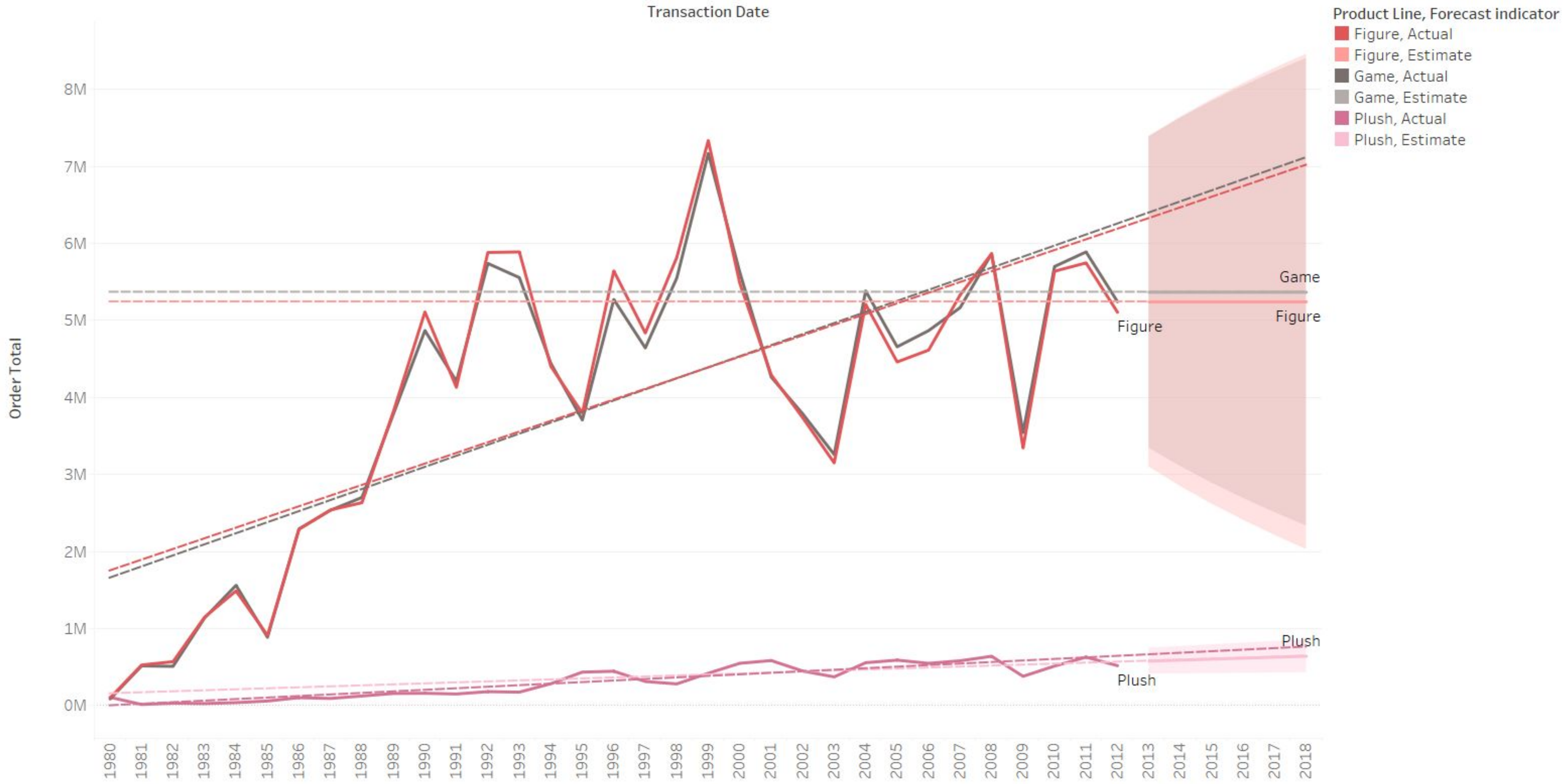
Novelty Brand	Actual	Estimate
Beach	Order Total = 118382*Year of Transaction Date + -2.34653e+08 R-Squared: 0.679727 P-value: < 0.0001	Order Total = 115768*Year of Transaction Date + -2.29945e+08 R-Squared: 1 P-value: < 0.0001
Bead	Order Total = 241704*Year of Transaction Date + -4.79188e+08 R-Squared: 0.725366 P-value: < 0.0001	Order Total = 248613*Year of Transaction Date + -4.94009e+08 R-Squared: 1 P-value: < 0.0001
Gift	Order Total = 152499*Year of Transaction Date + -3.02279e+08 R-Squared: 0.583032 P-value: < 0.0001	Order Total = 0*Year of Transaction Date + 3.77278e+06 R-Squared: 1 P-value: N/A
Promo	Order Total = 51775.8*Year of Transaction Date + -1.02133e+08 R-Squared: 0.249365 P-value: 0.0129784	Order Total = -161043*Year of Transaction Date + 3.26452e+08 R-Squared: 0.86218 P-value: 0.0714633
Thrift	Order Total = 2059.65*Year of Transaction Date + -3.88624e+06 R-Squared: 0.002729 P-value: 0.80845	Order Total = 0*Year of Transaction Date + 121504 R-Squared: 1 P-value: N/A

Toy - Product Brand (Time Series)



The trend of sum of Order Total for Transaction Date Year. Color shows details about Product Line. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Toy.

Toy - Product Brand (Time Series) Actual vs Estimation

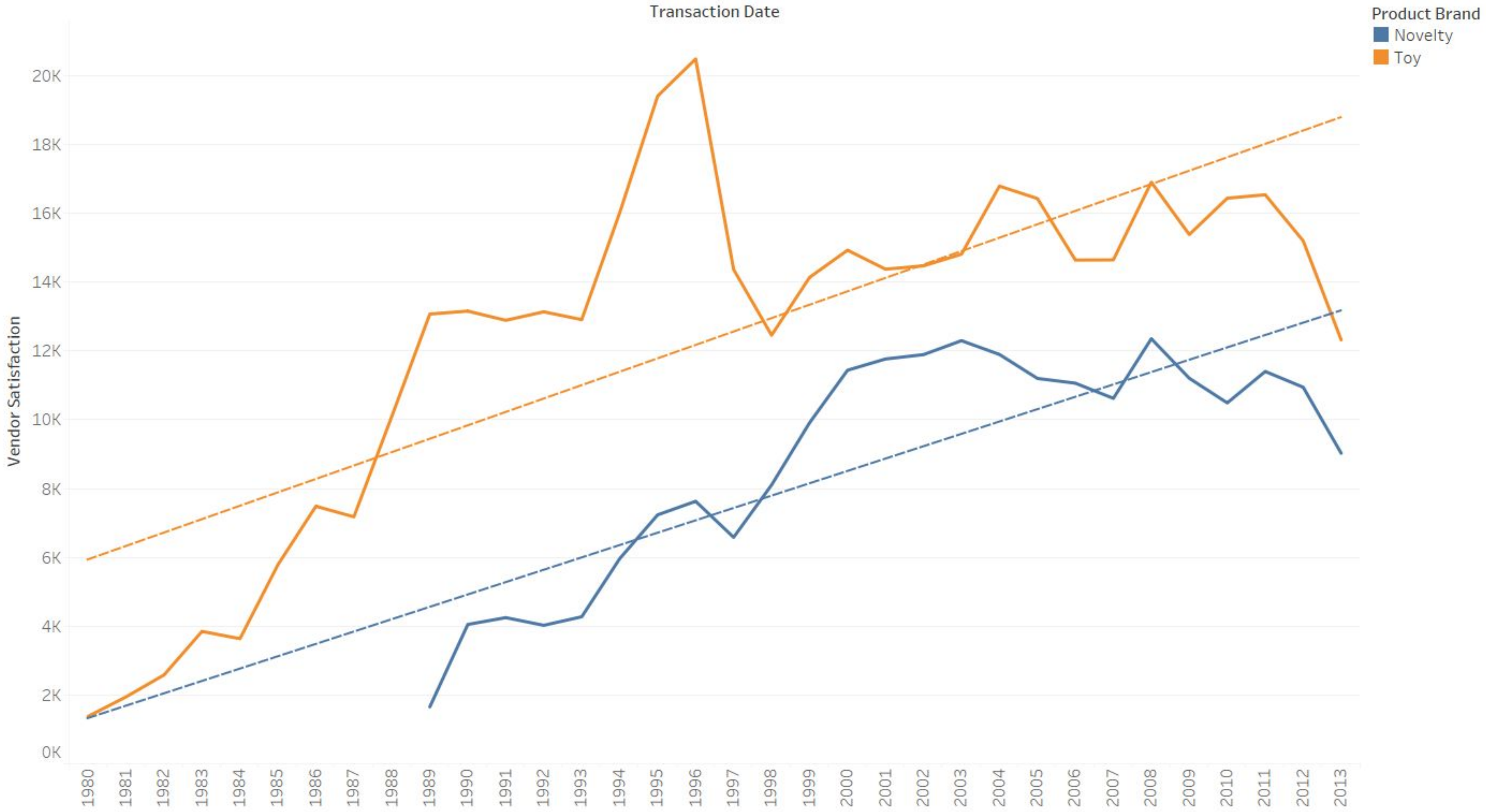


The trend of sum of Order Total (actual & forecast) for Transaction Date Year. Color shows details about Product Line and Forecast indicator. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Toy.

Toy Brand	Actual	Estimate
Figure	Order Total = 138633*Year of Transaction Date + -2.72747e+08 R-Squared: 0.503776 P-value: < 0.0001	Order Total = 0*Year of Transaction Date + 5.24089e+06 R-Squared: 1 P-value: N/A
Game	Order Total = 143656*Year of Transaction Date + -2.82786e+08 R-Squared: 0.55777 P-value: < 0.0001	Order Total = 0*Year of Transaction Date + 5.36574e+06 R-Squared: 1 P-value: N/A
Plush	Order Total = 20051.3*Year of Transaction Date + -3.9705e+07 R-Squared: 0.843818 P-value: < 0.0001	Order Total = 12831.7*Year of Transaction Date + -2.52542e+07 R-Squared: 1 P-value: < 0.0001

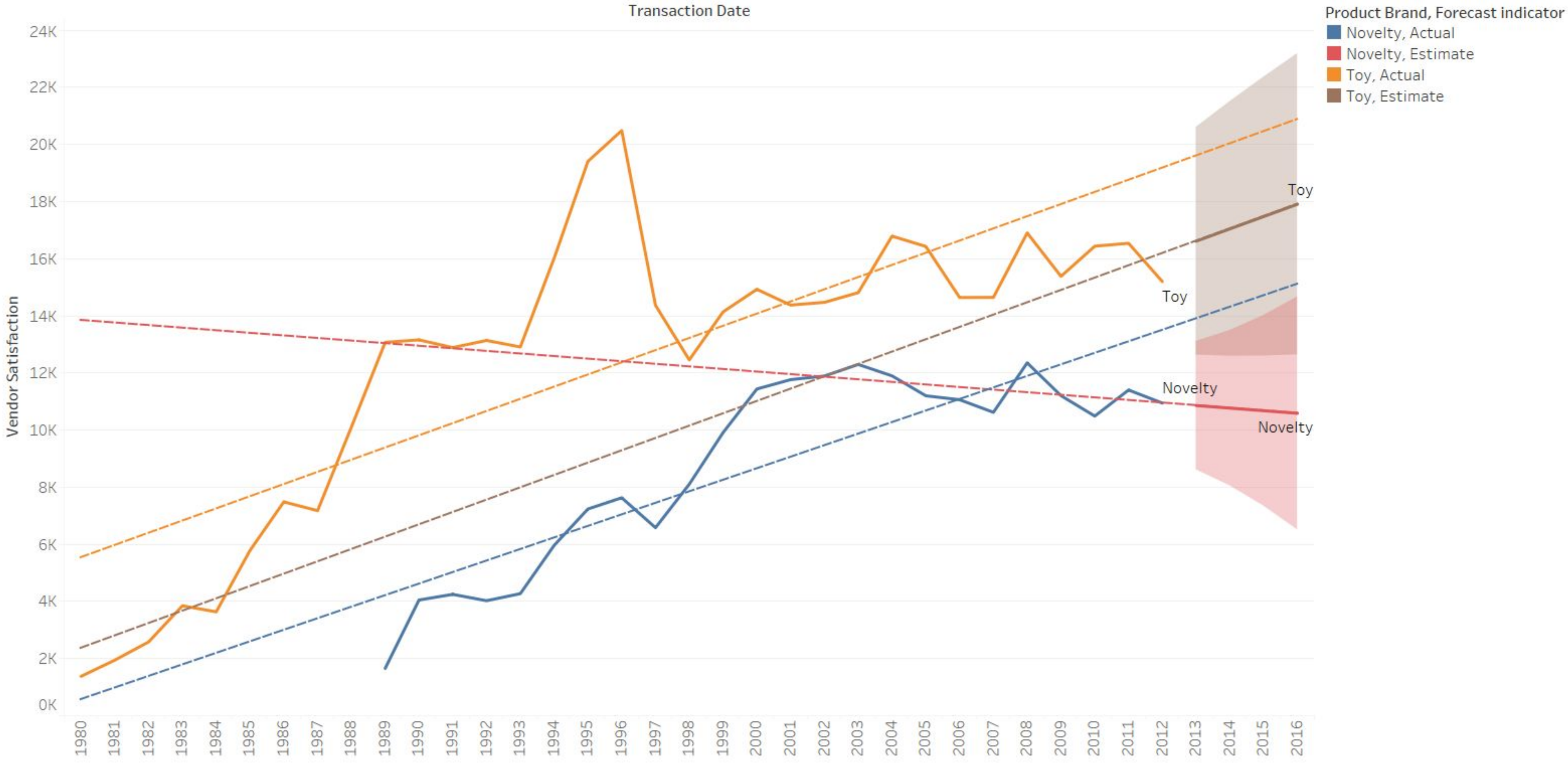
Product Understanding: Vendor Satisfaction

Vendor Satistfaction by Product Brand(Time Series)



The trend of sum of Vendor Satisfaction for Transaction Date Year. Color shows details about Product Brand.

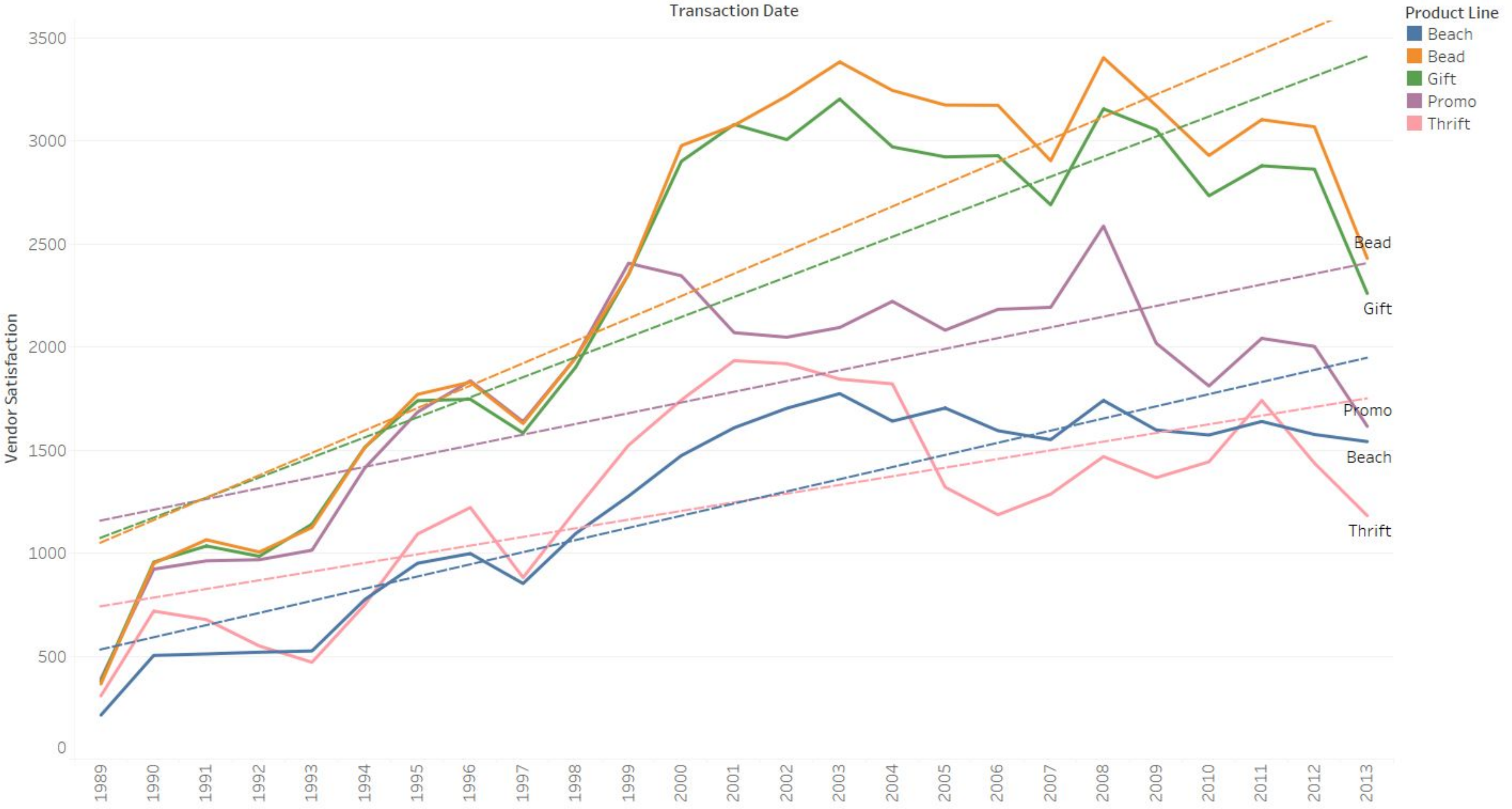
Vendor Satisfaction vs Product Brand (Time Series) Actual vs Estimate



The trend of sum of Vendor Satisfaction (actual & forecast) for Transaction Date Year. Color shows details about Product Brand and Forecast indicator. The marks are labeled by Product Brand.

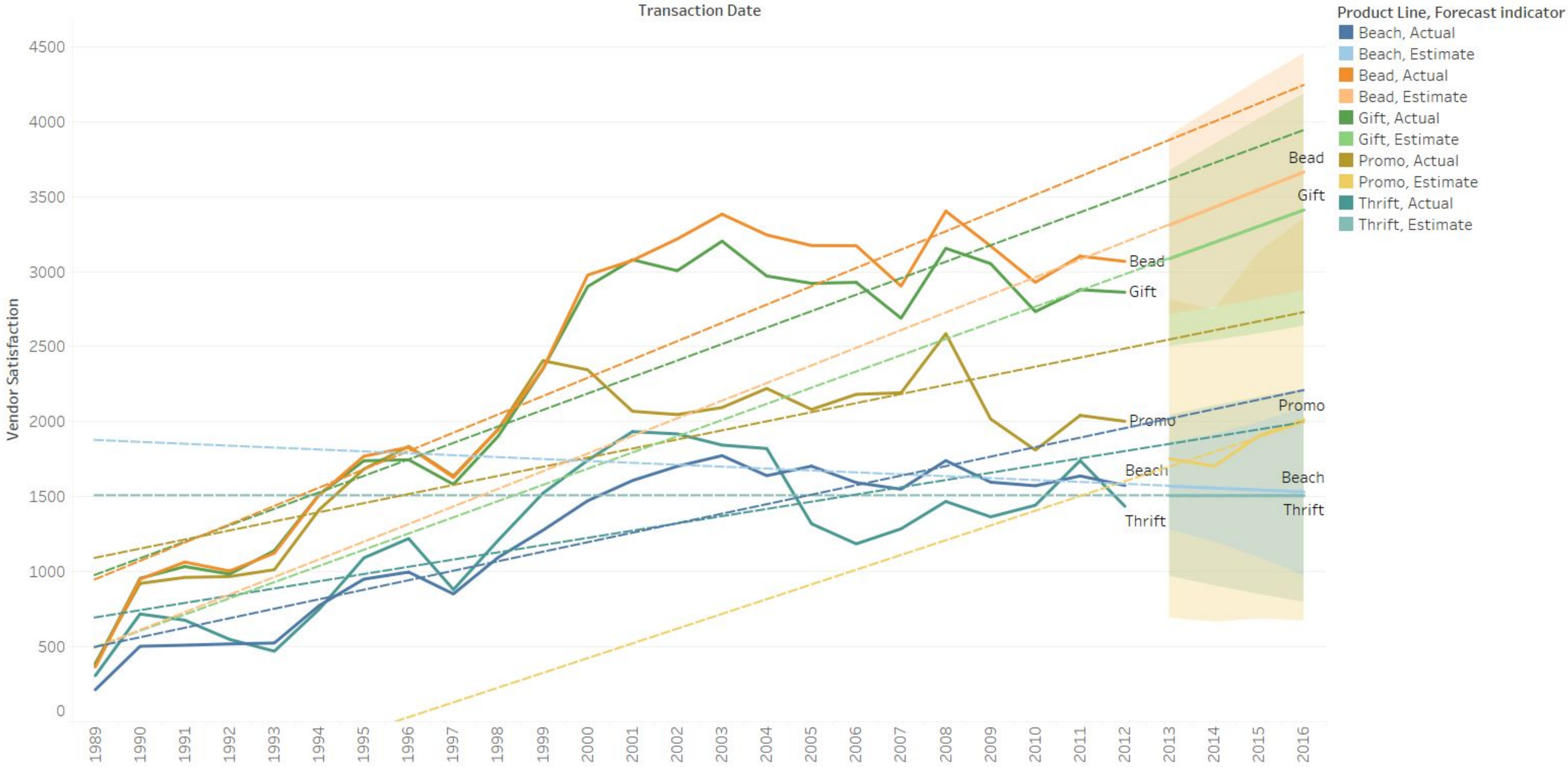
Product Brand	Real	Estimate
Novelty	Vendor Satisfaction = 403.974*Year of Transaction Date + -799306 R-Squared: 0.750511 P-value: < 0.0001	Vendor Satisfaction = -90.386*Year of Transaction Date + 192806 R-Squared: 1 P-value: < 0.0001
Toy	Vendor Satisfaction = 426.164*Year of Transaction Date + -838273 R-Squared: 0.639007 P-value: < 0.0001	Vendor Satisfaction = 431.876*Year of Transaction Date + -852755 R-Squared: 1 P-value: < 0.0001

Novelty Vendor Satisfaction - Product Brand (Time Series)



The trend of sum of Vendor Satisfaction for Transaction Date Year. Color shows details about Product Line. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Novelty.

Novelty Vendor Satisfaction - Product Brand - Real vs Estimate (Time Series)



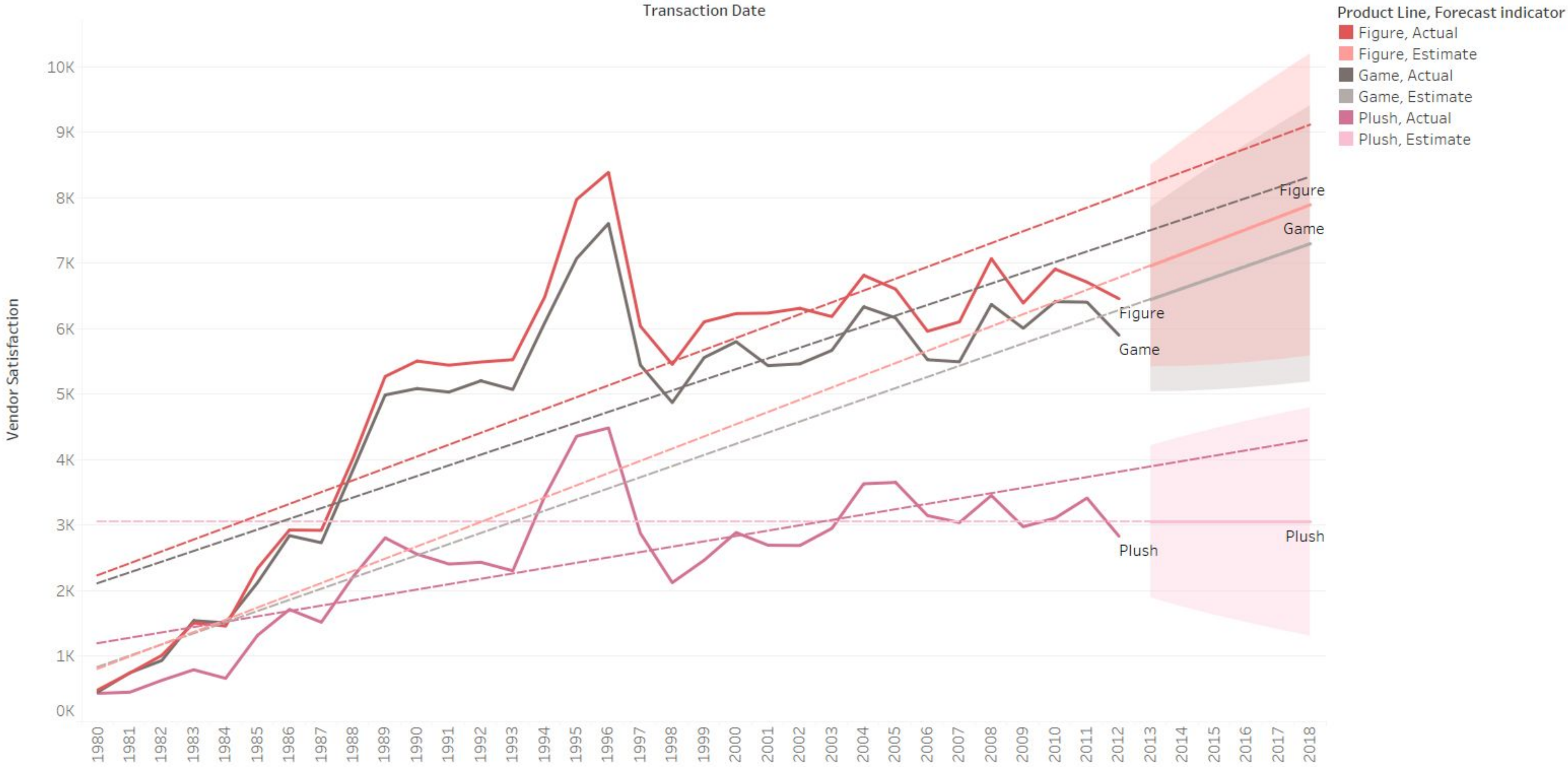
The trend of sum of Vendor Satisfaction (actual & forecast) for Transaction Date Year. Color shows details about Product Line and Forecast indicator. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Novelty.

Novelty Brand	Actual	Estimate
Beach	Vendor Satisfaction = 63.3629*Year of Transaction Date + -125533 R-Squared: 0.805779 P-value: < 0.0001	Vendor Satisfaction = -12.7174*Year of Transaction Date + 27170.1 R-Squared: 1 P-value: < 0.0001
Bead	Vendor Satisfaction = 122.008*Year of Transaction Date + -241728 R-Squared: 0.805257 P-value: < 0.0001	Vendor Satisfaction = 117.522*Year of Transaction Date + -233261 R-Squared: 1 P-value: < 0.0001
Gift	Vendor Satisfaction = 109.784*Year of Transaction Date + -217385 R-Squared: 0.778548 P-value: < 0.0001	Vendor Satisfaction = 108.01*Year of Transaction Date + -214338 R-Squared: 1 P-value: < 0.0001
Promo	Vendor Satisfaction = 60.624*Year of Transaction Date + -119492 R-Squared: 0.588194 P-value: < 0.0001	Vendor Satisfaction = 98.2426*Year of Transaction Date + -196065 R-Squared: 0.800328 P-value: 0.105389
Thrift	Vendor Satisfaction = 48.1953*Year of Transaction Date + -95168.5 R-Squared: 0.503377 P-value: 0.0001034	Vendor Satisfaction = 0*Year of Transaction Date + 1506.56 R-Squared: 1 P-value: N/A

Toy Vendor Satisfaction - Product Brand - (Time Series)

The trend of sum of Vendor Satisfaction for Transaction Date Year. Color shows details about Product Line. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Toy.

Toy Vendor Satisfaction - Product Brand - Real vs Estimate (Time Series)

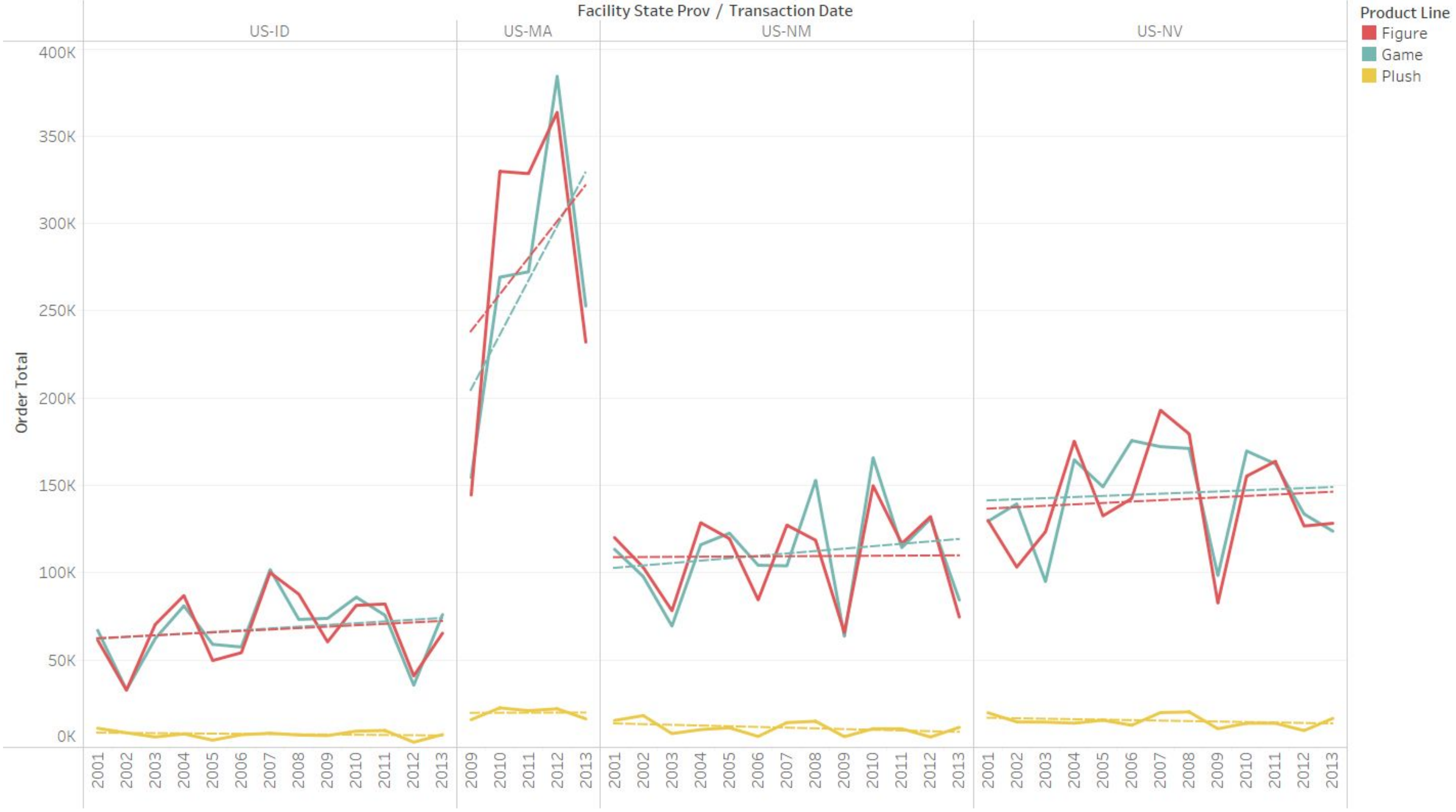


The trend of sum of Vendor Satisfaction (actual & forecast) for Transaction Date Year. Color shows details about Product Line and Forecast indicator. The marks are labeled by Product Line. The data is filtered on Product Brand, which keeps Toy.

Toy Brand	Actual	Estimate
Figure	Vendor Satisfaction = 181.041*Year of Transaction Date + -356231 R-Squared: 0.651709 P-value: < 0.0001	Vendor Satisfaction = 186.648*Year of Transaction Date + -368762 R-Squared: 1 P-value: < 0.0001
Game	Vendor Satisfaction = 163.336*Year of Transaction Date + -321296 R-Squared: 0.654605 P-value: < 0.0001	Vendor Satisfaction = 170.233*Year of Transaction Date + -336231 R-Squared: 1 P-value: < 0.0001
Plush	Vendor Satisfaction = 81.7875*Year of Transaction Date + -160747 R-Squared: 0.552237 P-value: < 0.0001	Vendor Satisfaction = 0*Year of Transaction Date + 3053.36 R-Squared: 1 P-value: N/A

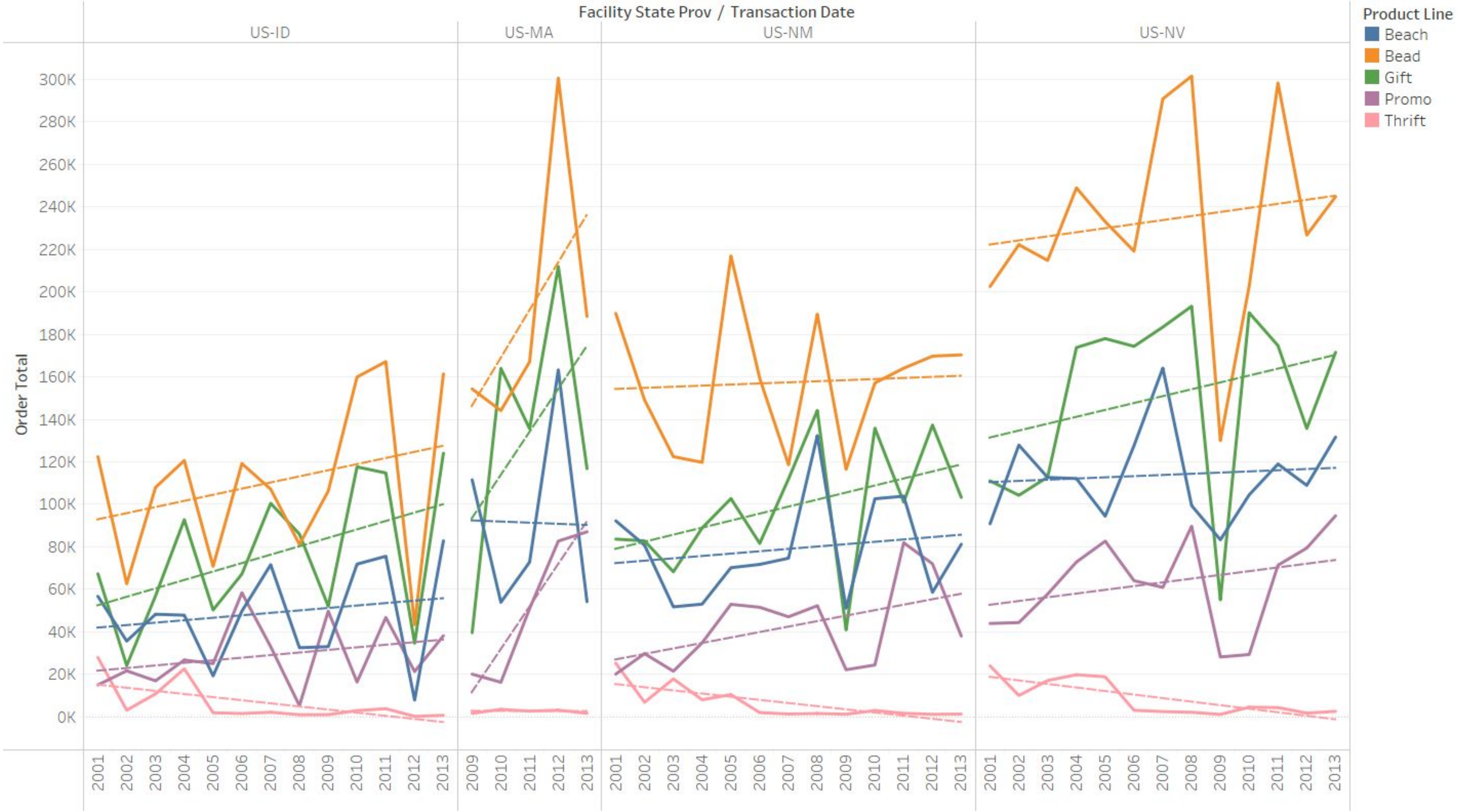
Product Revenue Understanding: Top performing States(Increasing vs Decreasing)

Upward trend (State) - Toy Line



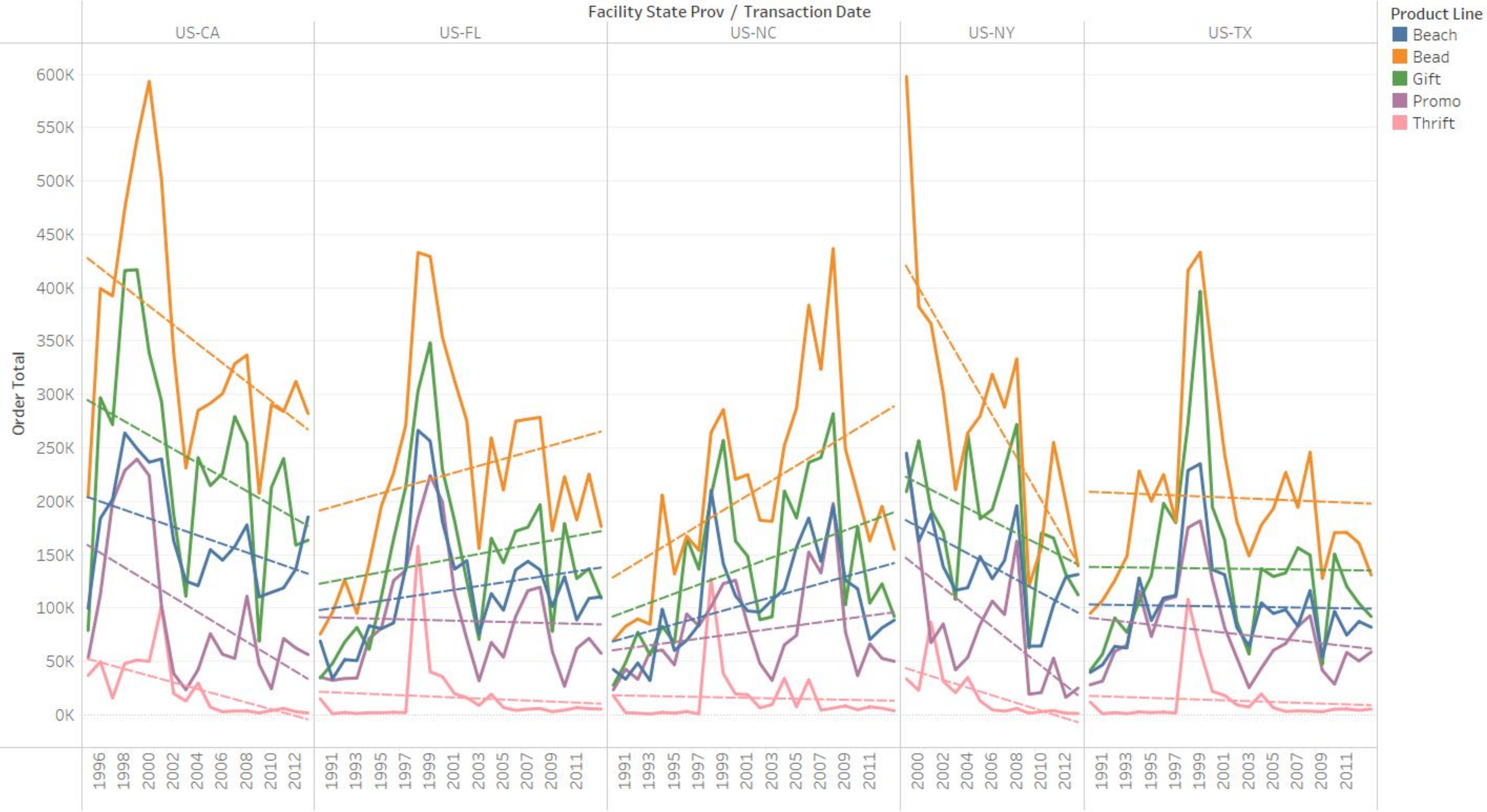
The trend of sum of Order Total for Transaction Date Year broken down by Facility State Prov. Color shows details about Product Line. The data is filtered on Product Brand, which keeps Toy. The view is filtered on Facility State Prov and Transaction Date Year. The Facility State Prov filter keeps US-ID, US-MA, US-NM and US-NV. The Transaction Date Year filter keeps 34 of 34 members.

Upward trend (State) - Novelty Line



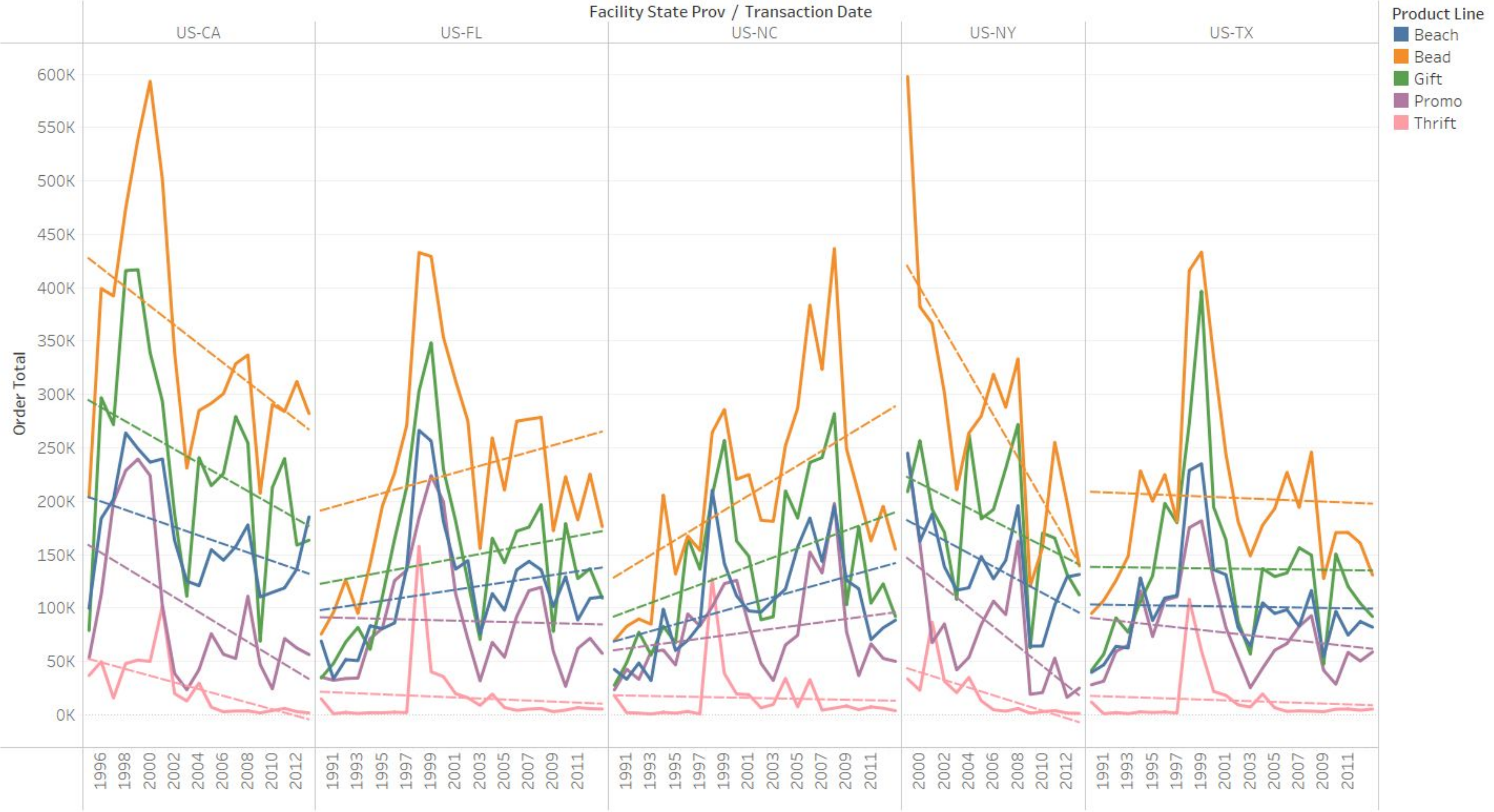
The trend of sum of Order Total for Transaction Date Year broken down by Facility State Prov. Color shows details about Product Line. The data is filtered on Product Brand, which keeps Novelty. The view is filtered on Facility State Prov and Transaction Date Year. The Facility State Prov filter keeps US-ID, US-MA, US-NM and US-NV. The Transaction Date Year filter keeps 34 of 34 members.

Downward trend(State) - Novelty Line



The trend of sum of Order Total for Transaction Date Year broken down by Facility State Prov. Color shows details about Product Line. The data is filtered on Product Brand, which keeps Novelty. The view is filtered on Facility State Prov and Transaction Date Year. The Facility State Prov filter keeps US-CA, US-FL, US-NC, US-NY and US-TX. The Transaction Date Year filter has multiple members selected.

Downward trend(State) - Toy Line



The trend of sum of Order Total for Transaction Date Year broken down by Facility State Prov. Color shows details about Product Line. The data is filtered on Product Brand, which keeps Novelty. The view is filtered on Facility State Prov and Transaction Date Year. The Facility State Prov filter keeps US-CA, US-FL, US-NC, US-NY and US-TX. The Transaction Date Year filter has multiple members selected.

Data Modeling using Multiple Linear Regression

Revenue Modeling

- Created 2 regression models, each one for states where revenue are increasing and decreasing
- The plan is to continuously improve what’s working and remove what’s not. We are also trying to figure out strategies from increasing states that can be implemented for decreasing states.
- We are only considering relevant independent variables which has p value smaller than 0.05 and r squared greater than 0.90

Regression Model(Model-D) For Decreasing Orders States

Regression Model For Decreasing Revenue States

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             OrderTotal   R-squared:                       0.990
Model:                            OLS   Adj. R-squared:                  0.990
Method:                 Least Squares   F-statistic:                 1.540e+06
Date:                Wed, 24 May 2023   Prob (F-statistic):               0.00
Time:                        10:58:18   Log-Likelihood:            -1.0125e+06
No. Observations:              175994   AIC:                         2.025e+06
Df Residuals:                  175982   BIC:                         2.025e+06
Df Model:                          11
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                   -55.6848      3.922    -14.198      0.000     -63.372     -47.998
VendorDistance            1.2713      0.032     39.475      0.000       1.208       1.334
MarketPenetration      -3.49e+05   1.01e+04    -34.396      0.000   -3.69e+05   -3.29e+05
OrderMarketingCost       -0.2890      0.315     -0.918      0.359      -0.906       0.328
OrderProductCost          2.6949      0.035     77.814      0.000       2.627       2.763
OrderDistributionCost     0.3439      0.008     42.589      0.000       0.328       0.360
OrderSalesCost            2.1689      0.247      8.777      0.000       1.685       2.653
SalesRepOrders           -0.6002      0.024    -25.084      0.000      -0.647      -0.553
SalesRepTarget            0.0023    3.8e-05     60.302      0.000       0.002       0.002
SalesRepActual            0.0006   5.73e-05      9.620      0.000       0.000       0.001
SalesRepVendors           0.3385      0.002    147.389      0.000       0.334       0.343
SalesRepRating          -55.8890      3.944    -14.171      0.000     -63.619     -48.159
VendorSatisfaction       60.7640      7.954      7.640      0.000      45.175      76.353
VendorRating              1.6992      0.598      2.841      0.004       0.527       2.871
==============================================================================
Omnibus:                   107135.385   Durbin-Watson:                   1.794
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4426593.605
Skew:                           2.323   Prob(JB):                         0.00
Kurtosis:                      27.126   Cond. No.                     2.42e+19
==============================================================================
```

- 0.99 R-squared - 99% of the dependent variable can be explained by the independent variables
- P-value is lower than 0.05 for most of the variables except for Order Marketing Cost
- Decided to run another model excluding Order Marketing Cost

Regression Model For Decreasing Revenue States

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             OrderTotal   R-squared:                       0.989
Model:                            OLS   Adj. R-squared:                  0.989
Method:                 Least Squares   F-statistic:                 2.070e+06
Date:                Wed, 24 May 2023   Prob (F-statistic):               0.00
Time:                        11:44:22   Log-Likelihood:            -1.0145e+06
No. Observations:              175994   AIC:                         2.029e+06
Df Residuals:                  175985   BIC:                         2.029e+06
Df Model:                           8
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.7788      2.020      0.386      0.700      -3.180       4.738
OrderDistributionCost   266.8265      0.067   3983.516      0.000     266.695     266.958
VendorDistance          -65.4388      0.041  -1597.067      0.000     -65.519     -65.359
MarketPenetration     -4.387e+05   1.01e+04    -43.252      0.000   -4.59e+05   -4.19e+05
SalesRepActual           -0.0005   3.56e-05    -13.764      0.000      -0.001      -0.000
SalesRepVendors           0.2855      0.002    135.271      0.000       0.281       0.290
SalesRepRating          -21.7874      6.241     -3.491      0.000     -34.020      -9.555
VendorRating              0.7190      0.605      1.189      0.234      -0.466       1.904
VendorSatisfaction       13.2749      8.010      1.657      0.097      -2.424      28.973
==============================================================================
Omnibus:                   106286.692   Durbin-Watson:                   1.786
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4151861.368
Skew:                           2.313   Prob(JB):                         0.00
Kurtosis:                      26.340   Cond. No.                     3.64e+08
==============================================================================
```

- 0.98 R-squared - 98% of the dependent variable can be explained by the independent variables
- P-value is lower than 0.05 for most of the variables except for some variables
- Decided to run another model excluding variables with P-value higher than 0.05. Also removed some variables for irrelevance and strong multicollinearity

Final Regression Model(Model-D) For Decreasing Revenue States

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             OrderTotal   R-squared:                       0.989
Model:                            OLS   Adj. R-squared:                  0.989
Method:                 Least Squares   F-statistic:                 2.759e+06
Date:                Wed, 24 May 2023   Prob (F-statistic):               0.00
Time:                        11:48:44   Log-Likelihood:            -1.0145e+06
No. Observations:              175994   AIC:                         2.029e+06
Df Residuals:                  175987   BIC:                         2.029e+06
Df Model:                           6
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     3.8963      1.120      3.479      0.001       1.701       6.091
OrderDistributionCost   266.8276      0.067   3984.766      0.000     266.696     266.959
VendorDistance          -65.4738      0.035  -1853.524      0.000     -65.543     -65.405
MarketPenetration     -4.384e+05   1.01e+04    -43.229      0.000   -4.58e+05   -4.19e+05
SalesRepActual           -0.0005   3.53e-05    -13.642      0.000      -0.001      -0.000
SalesRepVendors           0.2841      0.002    145.248      0.000       0.280       0.288
SalesRepRating          -11.6914      1.321     -8.849      0.000     -14.281      -9.102
==============================================================================
Omnibus:                106302.424   Durbin-Watson:                   1.786
Prob(Omnibus):               0.000   Jarque-Bera (JB):          4154475.599
Skew:                        2.314   Prob(JB):                         0.00
Kurtosis:                   26.348   Cond. No.                     3.64e+08
==============================================================================
```

- Good R-squared (98%)
- P-value for all the independent variables are lower than 0.05 which mean the model is statistically significant

YD=3.8+266.8*OrderDistributationCost-65.5*VendorDistance-438400*MarketPeneration-0.0005*SalesRepActual+0.28*SalesRepVendors-11.7*SalesRepRating

Regression Model (Model-I) For Increasing Orders States

Regression Model For Increasing Revenue States

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             OrderTotal   R-squared:                       0.977
Model:                            OLS   Adj. R-squared:                  0.977
Method:                 Least Squares   F-statistic:                 2.294e+05
Date:                Wed, 24 May 2023   Prob (F-statistic):               0.00
Time:                        11:56:27   Log-Likelihood:            -3.8698e+05
No. Observations:               59780   AIC:                         7.740e+05
Df Residuals:                   59768   BIC:                         7.741e+05
Df Model:                          11
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                  -568.0858     13.453    -42.227      0.000    -594.454    -541.718
VendorDistance            8.5505      0.103     83.243      0.000       8.349       8.752
MarketPenetration     -1.593e+06   3.55e+04    -44.878      0.000   -1.66e+06   -1.52e+06
OrderMarketingCost        0.1359      1.112      0.122      0.903      -2.044       2.315
OrderProductCost          2.5249      0.122     20.642      0.000       2.285       2.765
OrderDistributionCost     2.1608      0.026     83.861      0.000       2.110       2.211
OrderSalesCost           27.0783      0.862     31.403      0.000      25.388      28.768
SalesRepOrders           -3.4748      0.197    -17.617      0.000      -3.861      -3.088
SalesRepTarget            0.0002      0.000      1.158      0.247      -0.000       0.001
SalesRepActual            0.0065      0.000     35.770      0.000       0.006       0.007
SalesRepVendors           2.1883      0.019    116.904      0.000       2.152       2.225
SalesRepRating         -570.7082     13.530    -42.182      0.000    -597.227    -544.190
VendorSatisfaction     1102.0680     27.983     39.384      0.000    1047.222    1156.914
VendorRating             22.2907      2.156     10.337      0.000      18.064      26.517
==============================================================================
Omnibus:                    11091.413   Durbin-Watson:                   1.727
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           211027.550
Skew:                           0.360   Prob(JB):                         0.00
Kurtosis:                      12.176   Cond. No.                     2.12e+19
==============================================================================
```

- 0.97 R-squared - 97% of the dependent variable can be explained by the independent variables
- P-value is lower than 0.05 for most of the variables except for Order Marketing Cost and Sales Rep Target
- Decided to run another model excluding Order Marketing Cost and Sales Rep Target

Regression Model For Increasing Revenue States

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             OrderTotal   R-squared:                       0.977
Model:                            OLS   Adj. R-squared:                  0.977
Method:                 Least Squares   F-statistic:                 2.804e+05
Date:                Wed, 24 May 2023   Prob (F-statistic):               0.00
Time:                        11:57:49   Log-Likelihood:            -3.8698e+05
No. Observations:               59780   AIC:                         7.740e+05
Df Residuals:                   59770   BIC:                         7.741e+05
Df Model:                           9
Covariance Type:            nonrobust
===========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
const                  -568.1127     13.440    -42.270      0.000    -594.455    -541.770
VendorDistance            8.5440      0.102     83.397      0.000       8.343       8.745
MarketPenetration     -1.602e+06   3.46e+04    -46.312      0.000    -1.67e+06   -1.53e+06
OrderProductCost          2.5398      0.002   1423.400      0.000       2.536       2.543
OrderDistributionCost     2.1617      0.026     84.403      0.000       2.111       2.212
OrderSalesCost           27.1892      0.854     31.824      0.000      25.515      28.864
SalesRepOrders           -3.4973      0.196    -17.820      0.000      -3.882      -3.113
SalesRepActual            0.0065      0.000     36.383      0.000       0.006       0.007
SalesRepVendors           2.1880      0.019    116.901      0.000       2.151       2.225
SalesRepRating         -570.7461     13.517    -42.224      0.000    -597.240    -544.252
VendorSatisfaction     1104.9187     27.875     39.639      0.000    1050.284    1159.553
VendorRating             22.2635      2.156     10.325      0.000      18.037      26.490
==============================================================================
Omnibus:                    11095.154   Durbin-Watson:                   1.727
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           211179.928
Skew:                           0.360   Prob(JB):                         0.00
Kurtosis:                      12.180   Cond. No.                     1.55e+19
==============================================================================
```

- 0.97 R-squared - 97% of the dependent variable can be explained by the independent variables
- P-value is lower than 0.05 for all variables
- Decided to run another model excluding irrelevance and strong multicollinearity variables

Final Regression Model (Model-I) For Increasing Revenue States

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             OrderTotal   R-squared:                       0.976
Model:                            OLS   Adj. R-squared:                  0.976
Method:                 Least Squares   F-statistic:                 4.885e+05
Date:                Wed, 24 May 2023   Prob (F-statistic):               0.00
Time:                        12:01:24   Log-Likelihood:            -3.8794e+05
No. Observations:               59780   AIC:                         7.759e+05
Df Residuals:                   59774   BIC:                         7.759e+05
Df Model:                           5
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                   -75.9913      2.623    -28.976      0.000     -81.131     -70.851
OrderDistributionCost   257.0083      0.172   1497.846      0.000     256.672     257.345
VendorDistance          -54.8741      0.105   -521.745      0.000     -55.080     -54.668
MarketPenetration     -1.587e+06   3.48e+04    -45.575      0.000   -1.66e+06   -1.52e+06
SalesRepActual            0.0043      0.000     41.507      0.000       0.004       0.005
SalesRepVendors           2.0049      0.018    113.849      0.000       1.970       2.039
==============================================================================
Omnibus:                11296.142   Durbin-Watson:                   1.702
Prob(Omnibus):              0.000   Jarque-Bera (JB):           208375.351
Skew:                       0.397   Prob(JB):                         0.00
Kurtosis:                  12.112   Cond. No.                     4.53e+08
==============================================================================
```

- Good R-squared (97%)
- P-value for all the independent variables are lover than 0.05 which mean the model is statistically significant

YI= -75.9+257*OrderDistributationCost-54.9*VendorDistance-1587000*MarketPeneration+0.0043*SalesRepActual+2*SalesRepVendors

Model-D Vs Model-I

- Both models have same dependent variable(Y) as OrderTotal

- Almost same independent variables(X) with Model-I has one less variable

- Both model have high r-squared and statistically significant p-values

- The company should use the strategics, procedures, and systems that have been used in Model-I’s states to increase revenue and customer satisfaction in Model-D’s states as Model-I’s states revenue have been increased while using the same X as Model-D

Customer Satisfaction Modeling

- Created 1 regression model, using the segmented data sets(NA region, 1997+)
- The plan is to continuously improve what's working and remove what's not.
- We are only considering relevant independent variables which has p value smaller than 0.05 and r squared greater than 0.90

Regression Model For Customer Satisfaction

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     VendorSatisfaction   R-squared:                       0.958
Model:                            OLS   Adj. R-squared:                  0.958
Method:                 Least Squares   F-statistic:                 1.762e+06
Date:                Thu, 25 May 2023   Prob (F-statistic):               0.00
Time:                        20:00:24   Log-Likelihood:             1.8026e+06
No. Observations:              772181   AIC:                        -3.605e+06
Df Residuals:                  772170   BIC:                        -3.605e+06
Df Model:                          10
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.4752   8.18e-05   5809.832      0.000       0.475       0.475
VendorDistance           0.0011   3.47e-06    311.409      0.000       0.001       0.001
MarketPenetration       63.3703      0.573    110.538      0.000      62.247      64.494
OrderMarketingCost   -4.023e-06   4.62e-05     -0.087      0.931    -9.46e-05    8.66e-05
OrderProductCost      2.181e-05   5.11e-06      4.266      0.000     1.18e-05    3.18e-05
OrderDistributionCost    0.0003   8.72e-07    310.132      0.000       0.000       0.000
OrderSalesCost          -0.0277   1.47e-05  -1888.814      0.000      -0.028      -0.028
OrderTotal           -1.019e-05    2.2e-07    -46.327      0.000    -1.06e-05   -9.76e-06
SalesRepTarget          1.19e-08   6.01e-09      1.980      0.048     1.23e-10    2.37e-08
SalesRepVendors      -7.334e-05   3.59e-07   -204.396      0.000     -7.4e-05   -7.26e-05
SalesRepRating           0.4780   8.22e-05   5812.028      0.000       0.478       0.478
VendorType              -0.0006   1.78e-05    -35.516      0.000      -0.001      -0.001
VendorRating             0.0015   9.05e-05     16.746      0.000       0.001       0.002
==============================================================================
Omnibus:                   199051.601   Durbin-Watson:                   1.918
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          3303809.388
Skew:                           0.805   Prob(JB):                         0.00
Kurtosis:                      13.005   Cond. No.                     2.31e+19
==============================================================================
```

- 0.96 R-squared: 96% of the dependent variable can be explained by the independent variables
- P-value is lower than 0.05 for most of the variables except for Order Marketing Cost
- Decided to run another model excluding Order Marketing Cost

Regression Model For Customer Satisfaction

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     VendorSatisfaction   R-squared:                       0.958
Model:                            OLS   Adj. R-squared:                  0.958
Method:                 Least Squares   F-statistic:                 1.958e+06
Date:                Thu, 25 May 2023   Prob (F-statistic):               0.00
Time:                        20:00:25   Log-Likelihood:             1.8026e+06
No. Observations:              772181   AIC:                        -3.605e+06
Df Residuals:                  772171   BIC:                        -3.605e+06
Df Model:                           9
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.4752   7.84e-05   6060.112      0.000       0.475       0.475
VendorDistance            0.0011   3.46e-06    311.961      0.000       0.001       0.001
MarketPenetration        63.3702      0.573    110.538      0.000      62.247      64.494
OrderProductCost       2.137e-05   5.41e-07     39.486      0.000    2.03e-05    2.24e-05
OrderDistributionCost     0.0003   8.66e-07    312.263      0.000       0.000       0.000
OrderSalesCost           -0.0277   1.45e-05  -1912.758      0.000      -0.028      -0.028
OrderTotal            -1.019e-05    2.2e-07    -46.327      0.000   -1.06e-05   -9.76e-06
SalesRepTarget          1.19e-08   6.01e-09      1.980      0.048    1.23e-10    2.37e-08
SalesRepVendors       -7.334e-05   3.59e-07   -204.396      0.000    -7.4e-05   -7.26e-05
SalesRepRating            0.4780   7.89e-05   6054.249      0.000       0.478       0.478
VendorType               -0.0006   1.78e-05    -35.515      0.000      -0.001      -0.001
VendorRating              0.0015   9.05e-05     16.746      0.000       0.001       0.002
==============================================================================
Omnibus:                   199051.864   Durbin-Watson:                   1.918
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          3303821.817
Skew:                           0.805   Prob(JB):                         0.00
Kurtosis:                      13.005   Cond. No.                     2.31e+19
==============================================================================

Notes:
```

- Good R-squared (97%)
- P-value for all the independent variables are lover than 0.05 which mean the model is statistically significant

YC= 0.47 + 0.0011*VendorDistance + 63.3702*MarketPeneration + 0.00002*OrderProductCost + 0.0003*OrderDistributationCost -0.0277*OrderSalesCost - 0.00001*Ordertotal + 1.19e-08*SalesRepTarget - 7.334e-5*SalesRepVendors + 0.47*SalesRepRating - 0.0006*VendorType + 0.0015*VendorRating

Increasing the Revenue and Customer Satisfaction

- **Revenue**
  - **Revenue Decreasing States**

    Y=3.8+266.8*OrderDistributationCost-65.5*VendorDistance-438400*MarketPeneration-0.0005*SalesRepActual+0.28*SalesRepVendors-11.7*SalesRepRating
    Inorder to increase the revenue for decreasing state, we need to
    - Increase Order Distribution Cost
    - Reduce Vendor Distance
    - Reduce Market Penetration
    - Reduce SalesRep Actual
    - Increase SalesRepVendors
    - Reduce SalesRepRating

  - **Revenue Increasing State**

    Y= -75.9+257*OrderDistributationCost-54.9*VendorDistance-1587000*MarketPeneration+0.0043*SalesRepActual+2*SalesRepVendors
    Inorder to increase the revenue for increasing state, we need to
    - Increase Order Distribution Cost
    - Reduce Vendor Distance
    - Reduce Market Penetration
    - Increase SalesRep Actual
    - Increase SalesRepVendors
    - Reduce SalesRepRating

Increasing the Revenue and Customer Satisfaction(Contd)

- **Customer(Vendor) Satisfaction**
  YC= 0.47 + 0.0011*VendorDistance + 63.3702*MarketPeneration + 0.00002*OrderProductCost + 0.0003*OrderDistributationCost -0.0277*OrderSalesCost - 0.00001*Ordertotal + 1.19e-08*SalesRepTarget - 7.334e-5*SalesRepVendors + 0.47*SalesRepRating - 0.0006*VendorType + 0.0015*VendorRating
  Inorder to increase the revenue for increasing state, we need to
    - Increase VendorDistance
    - Increase Market Penetration
    - Increase OrderProductCost
    - Increase OrderDistributionCost
    - Reduce OrderSalesCost
    - Reduce Ordertotal
    - Increase SalesRepTarget
    - Reduce SalesRepVendors
    - Increase SalesRepRating
    - Reduce VendorType
    - Increase VendorRating

Deployment Tactics(Points to keep in mind)

- The decision to increase or decrease a variable should be determined by the coefficient. Higher the coefficient, higher will be the impact for the outcome. Eg in Y= 0.47 + 0.0011*VendorDistance + 63.3702*MarketPeneration, increasing the marketing penetration by 10% has higher impact than increasing the vendor distance by 10%
- All the three models ie YD, YC and YI has overlapping variables. Increasing one variable may increase the value for one model but decrease the value for another model. So, the business should prioiritize improving those variables that is in all or most of the model and has higher impact compared to other variables.
- If the above strategy doesn't work, then the business should prioritize improving one factor first ie revenue or customer satisfaction and then move on to improve another factor. This would be the optimum utilization of the resources.

Recommendations

- For both Revenue and Customer Satisfaction
  - Increase Order Distribution Cost: This indicates further study needs to be done regarding the supply chain of the business.
- For Revenue based on priority
  - Decrease Market Penetration
  - Decrease Vendor Distance
- For Customer Satisfaction (based on Priority)
  - Increase Market Penetration

Additional Recommendations for business

- Provide vendor loyalty program to all the vendors
- Develop marketing models and business strategies based on the digital outlets available and big data
- Do market research and develop models based on consumer psychology.
- Promote more Novelty(Beach) products.
- Investigate Toy(Figure and Game) more as they are the one that has taken significant hit over the year.

References

CRISP-DM Help Overview. (2021, August 17). IBM.

https://www.ibm.com/docs/it/spss-modeler/saas?topic=dm-crisp-help-overview

dot-com bubble. (n.d.). Britannica. https://www.britannica.com/event/dot-com-bubble

Time Series Analysis. (n.d.). Tableau. https://www.tableau.com/solutions/time-series-analysis