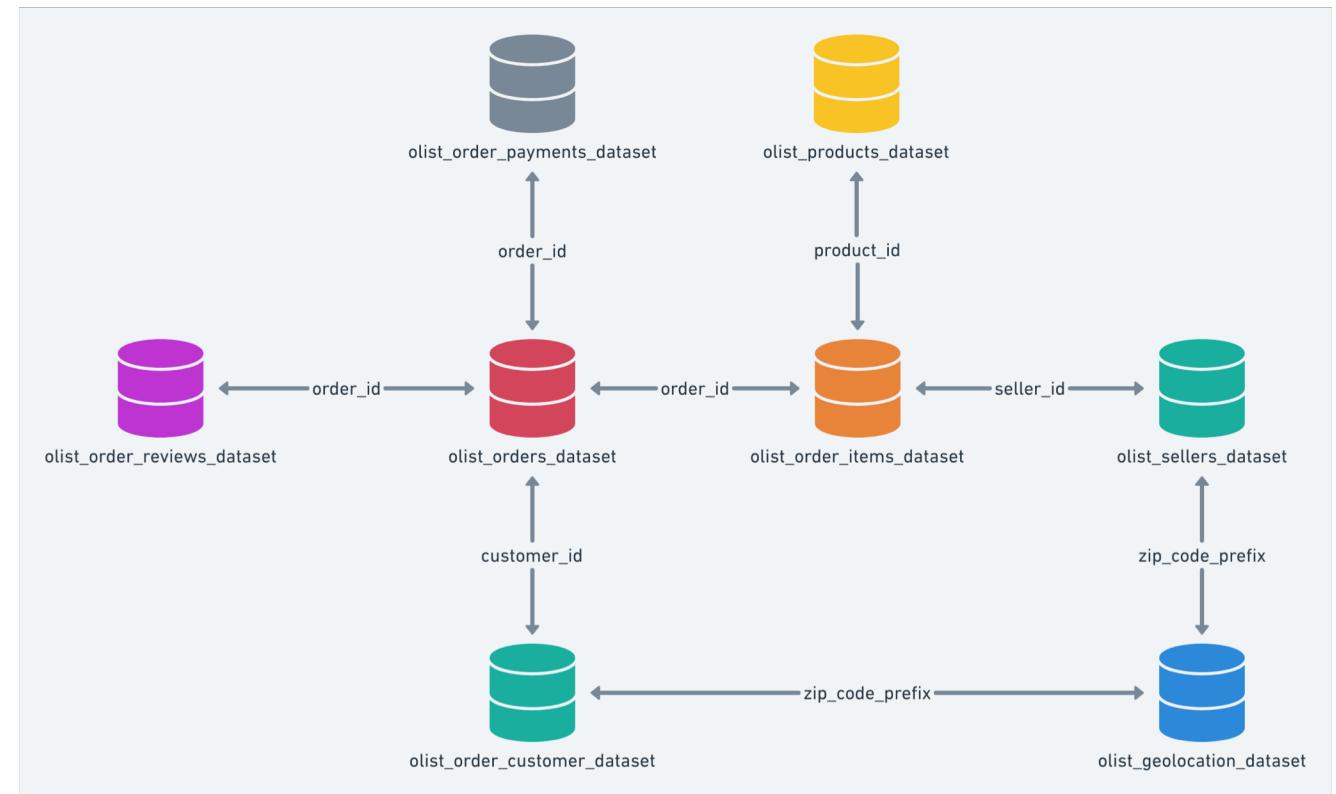


Consumer Behavior Analytics

Francis Troy Kirinhakone
Carlos He He

Dataset

- Brazilian E-commerce Public Dataset by Olist
- Contains over 100,000 orders with product, customer, and reviews info.



Description

*

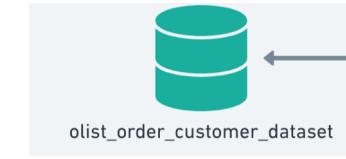
- Brazilian E-commerce Public Dataset by Olist
- Contains over 100,000 orders with product, customer, and reviews info.

Goal and Mindset

- Brazilian E-commerce Public Dataset by Olist
- Contains over 100,000 orders with product, customer, and reviews info.

*

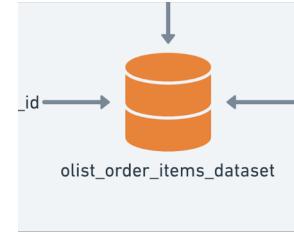
Dataset



```
```{r}
names(customers)
```

[1] "customer_id"           "customer_unique_id"      "customer_zip_code_prefix"
[4] "customer_city"          "customer_state"
```

Dataset

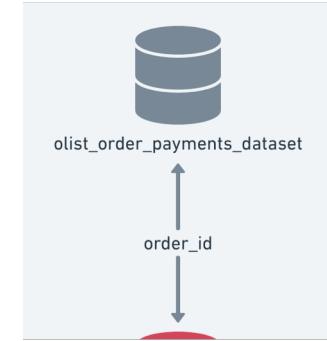


```
names(order_items)
```

```
[1] "order_id"      "order_item_id" "product_id"      "seller_id"      "price"  
[6] "freight_value"
```

Dataset

- Brazilian E-commerce Public Dataset by Olist
- Contains over 100,000 orders with product, customer, and reviews info.



```
{r}
order_payments %>%
  names()
```
[1] "order_id" "payment_sequential" "payment_type"
[4] "payment_installments" "payment_value"
```

# Dataset

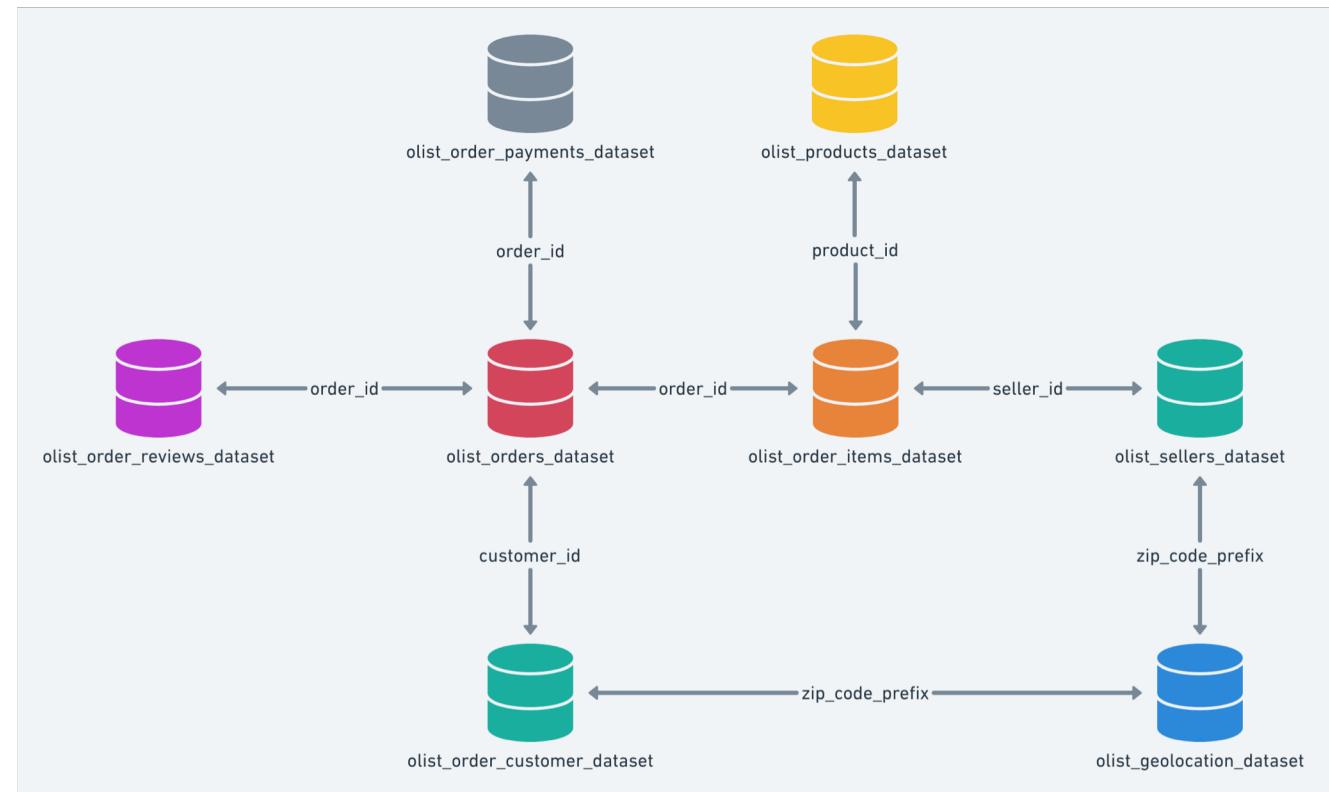
- Brazilian E-commerce Public Dataset by Olist
- Contains over 100,000 orders with product, customer, and reviews info.



```
{r}
orders %>% names()
```
[1] "order_id"                      "customer_id"
[3] "order_status"                   "order_purchase_timestamp"
[5] "order_approved_at"              "order_delivered_carrier_date"
[7] "order_delivered_customer_date"  "order_estimated_delivery_date"
```

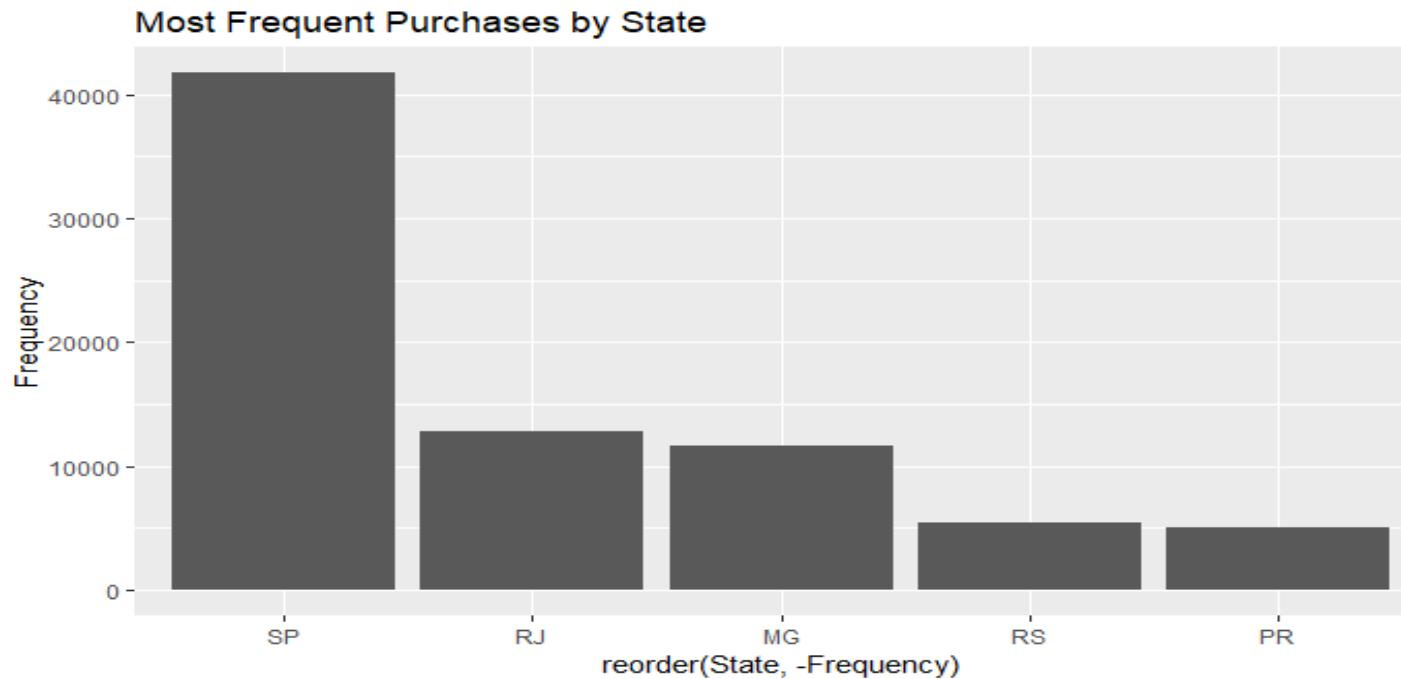
Dataset

- Brazilian E-commerce Public Dataset by Olist
 - Contains over 100,000 orders with product, customer, and reviews info.



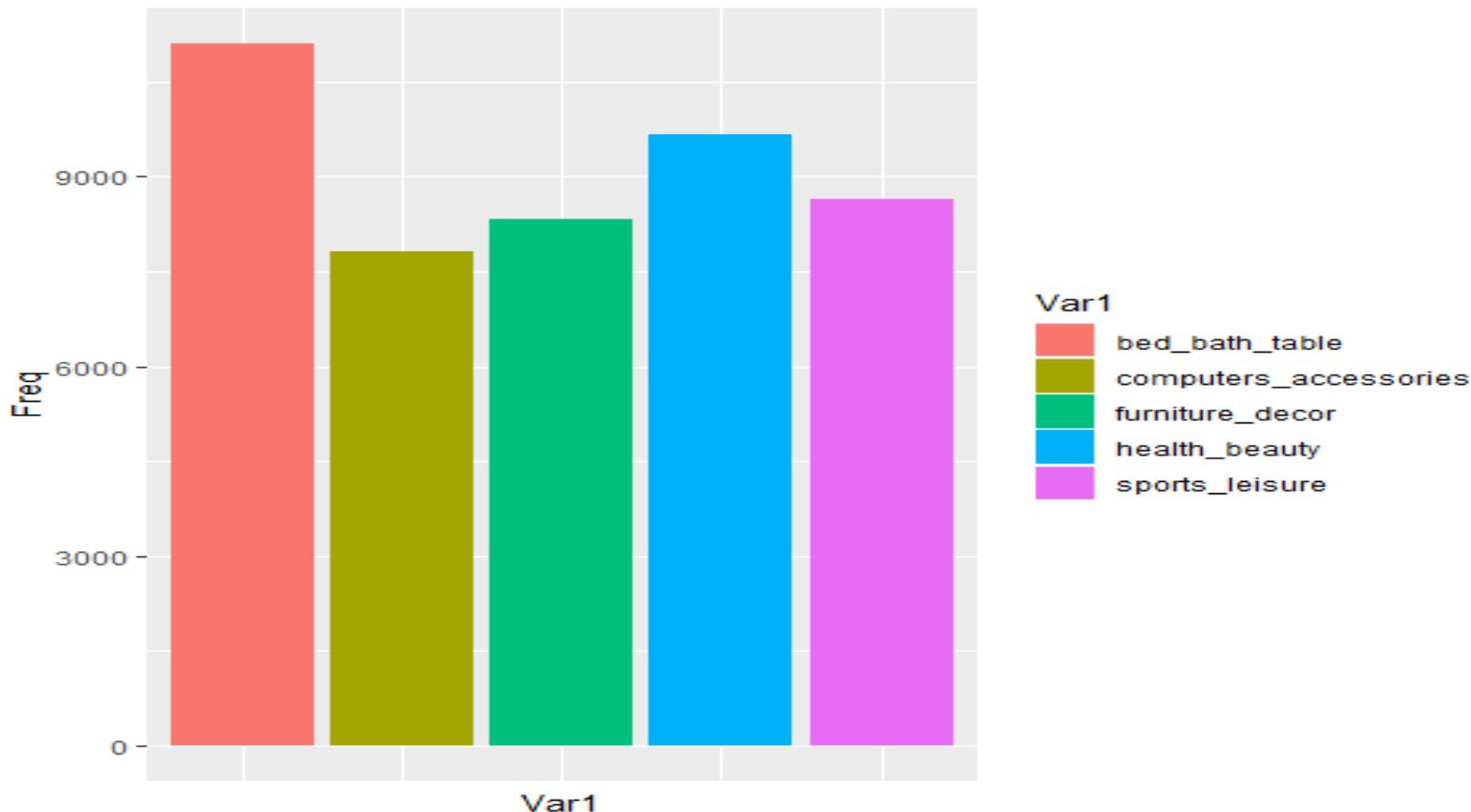
Exploratory Data Analysis

- Discovered the most frequent purchases by states.



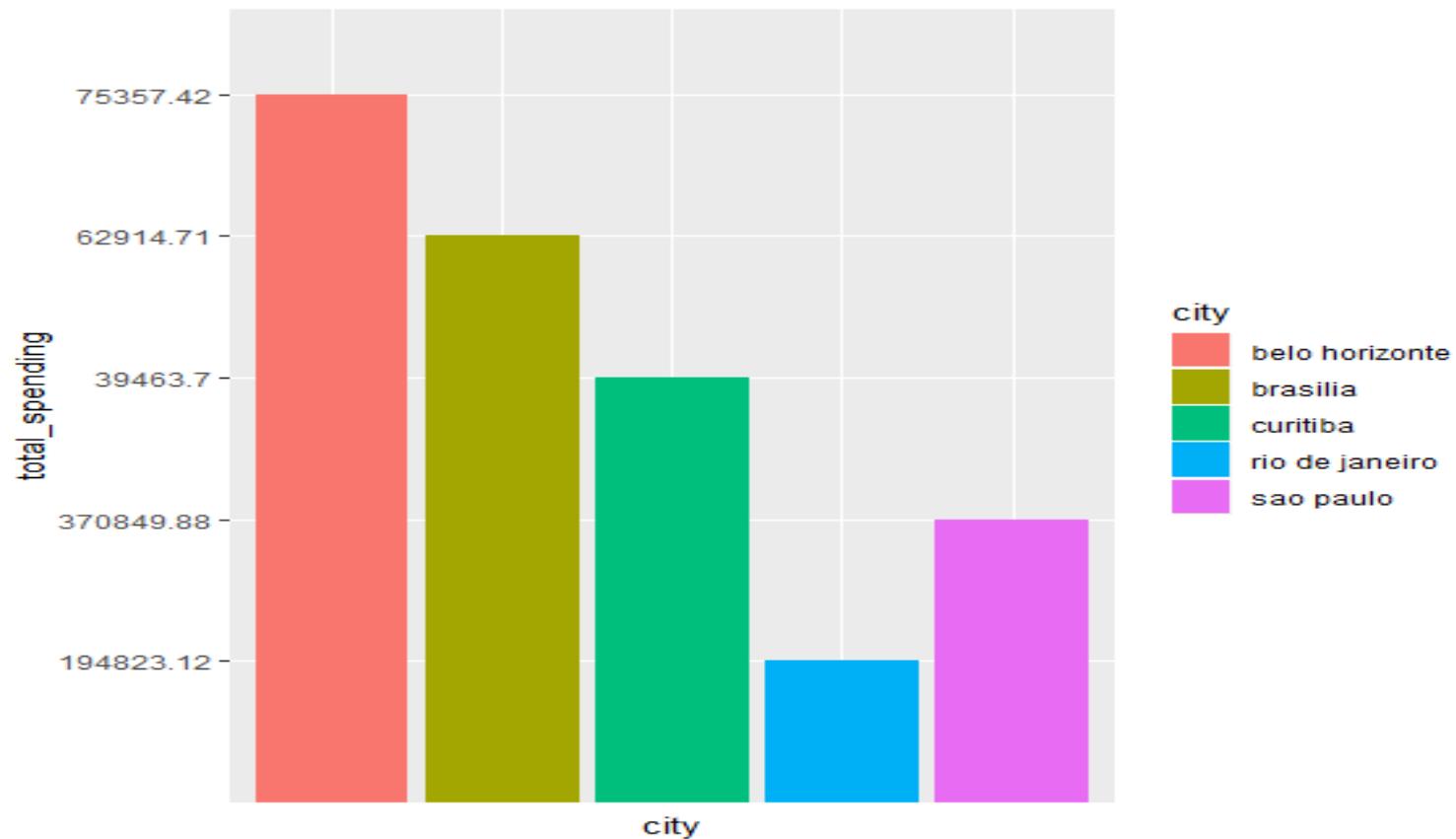
Exploratory Data Analysis

- Discovered most frequent bought products by category.



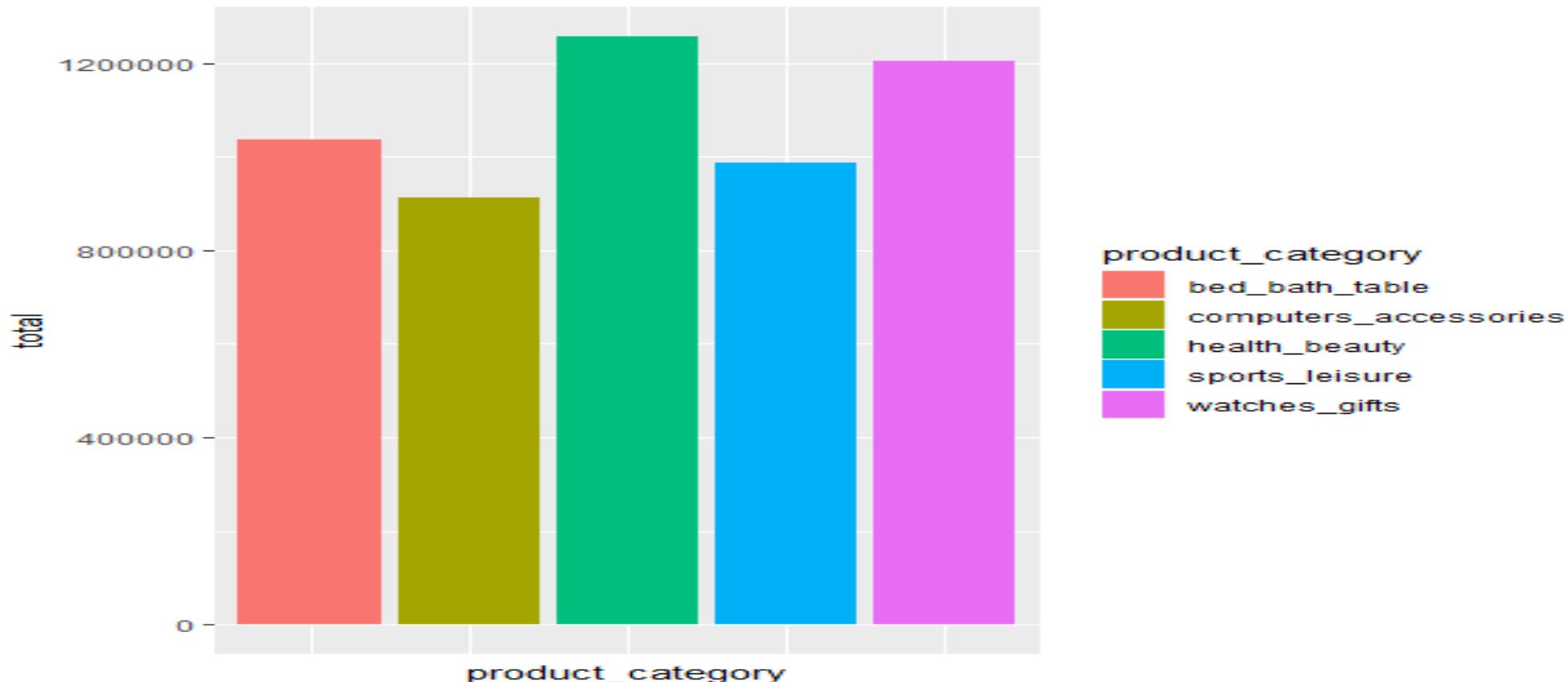
Exploratory Data Analysis

- Total spending of each city.

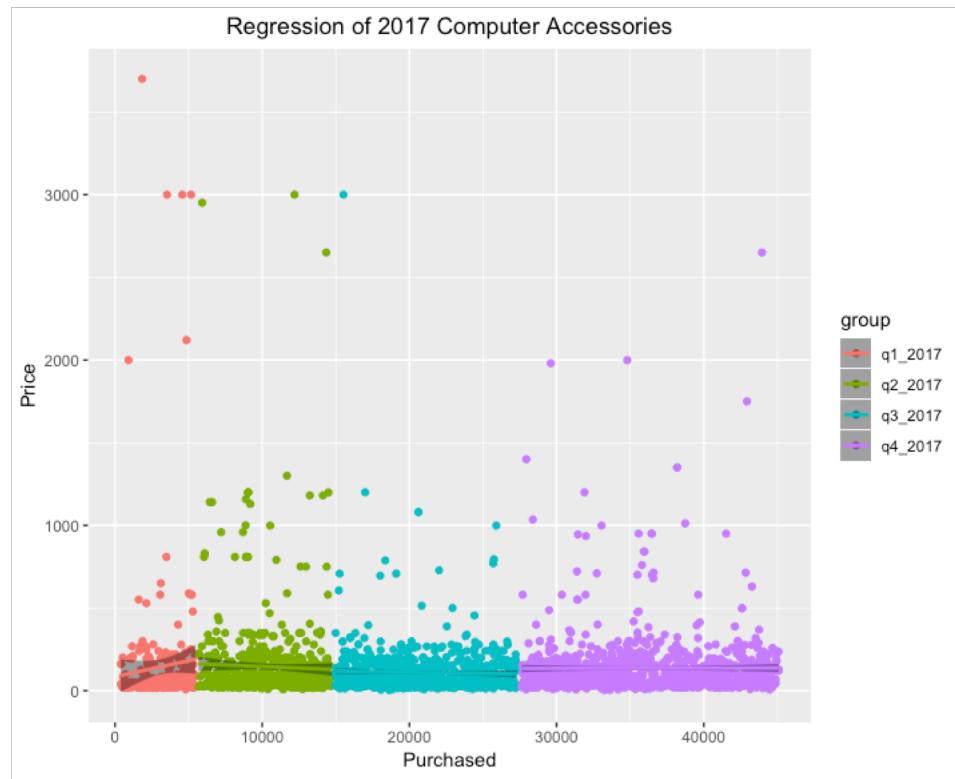


Exploratory Data Analysis

- Total spending for each category.



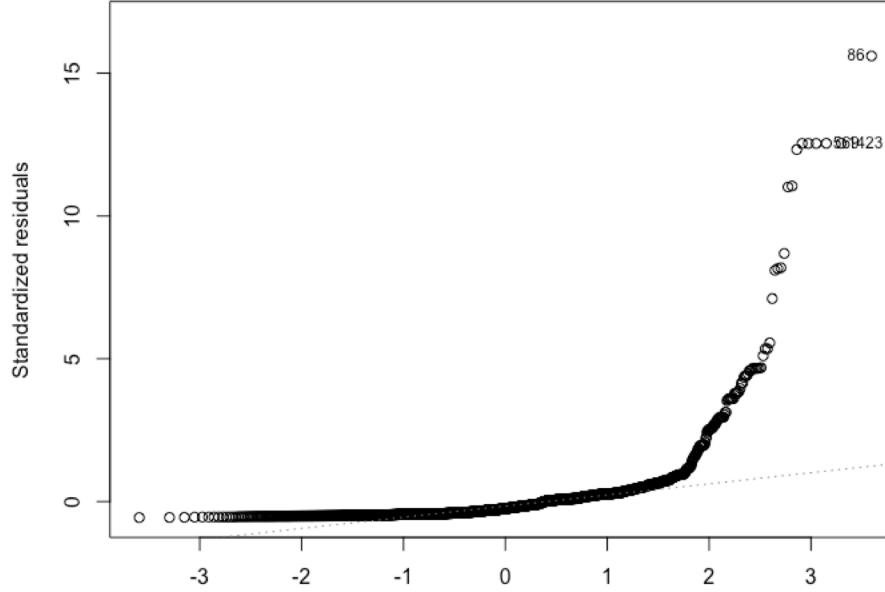
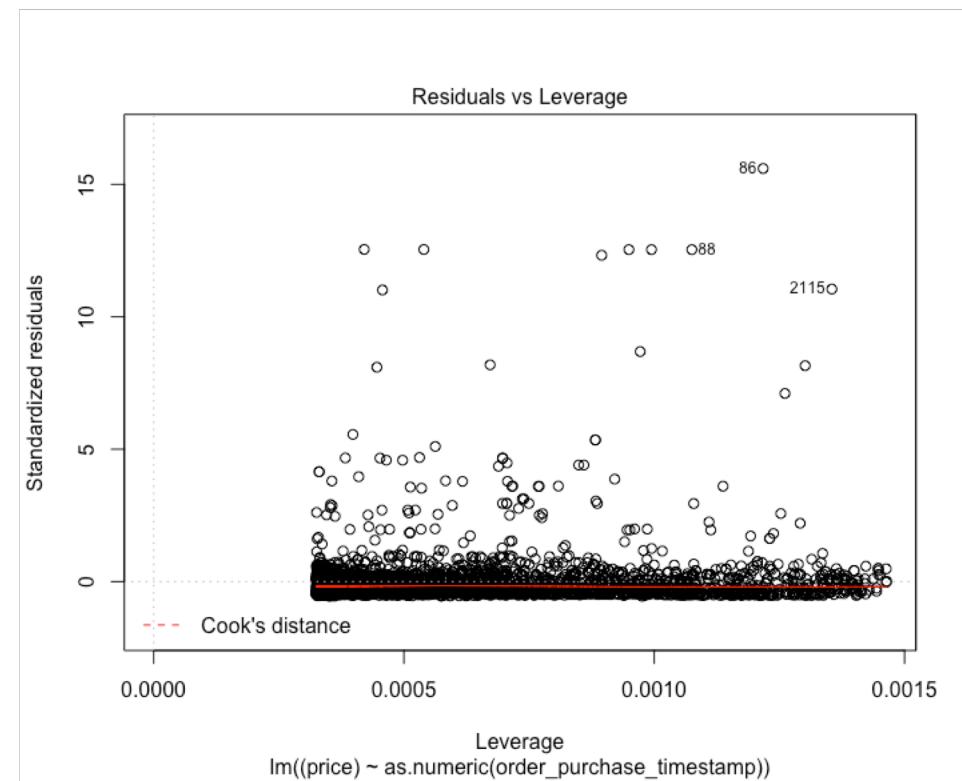
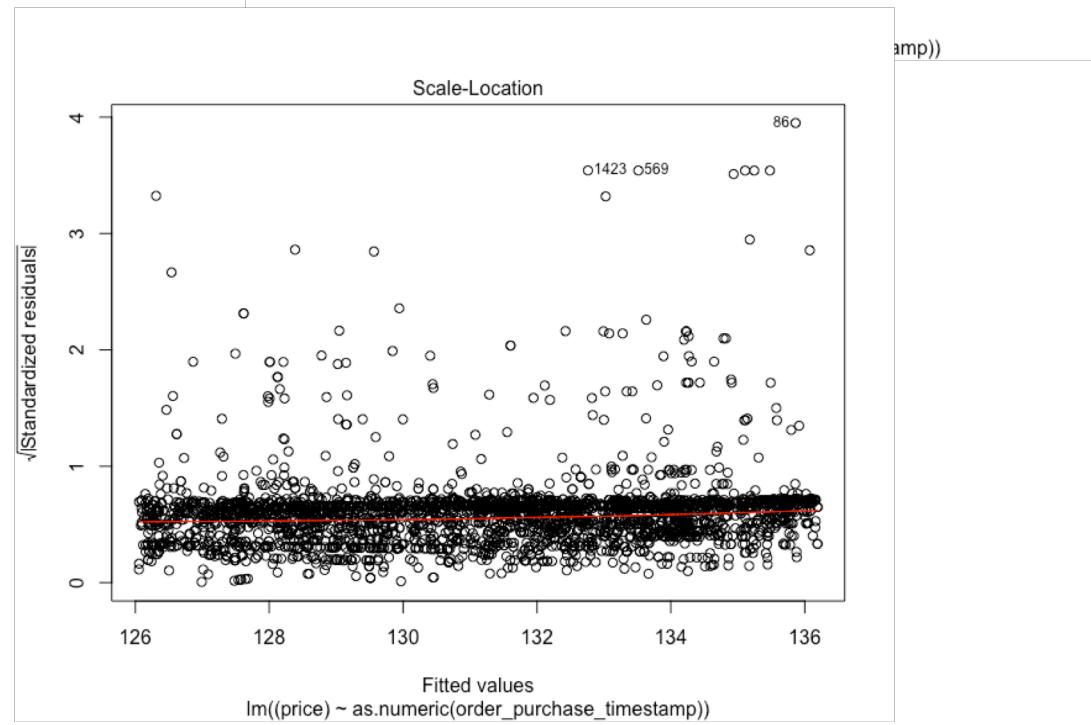
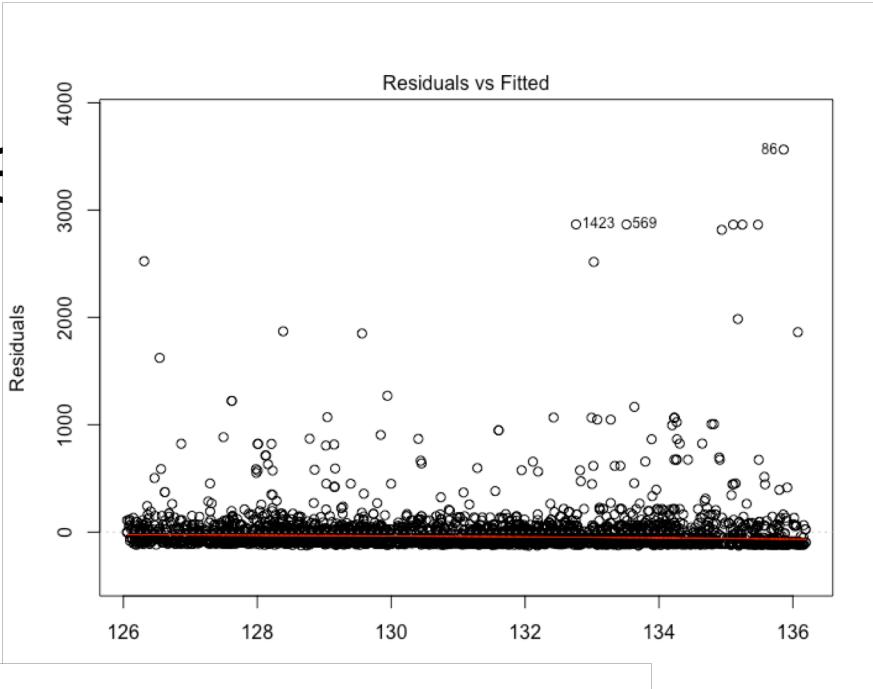
Linear Regression



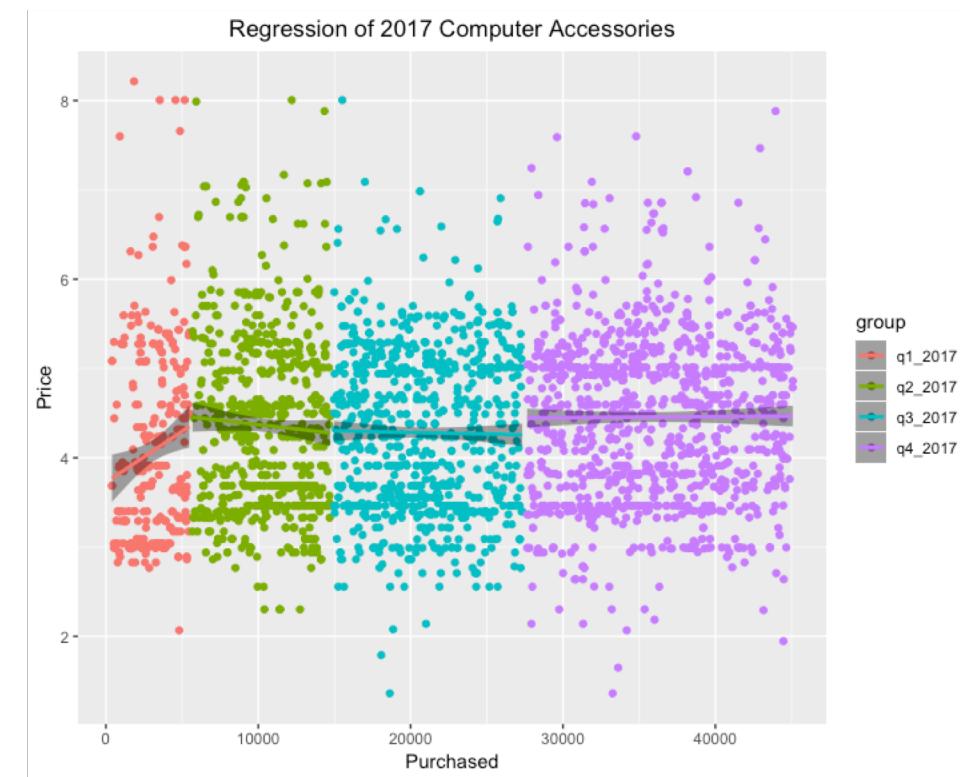
Linear Regression

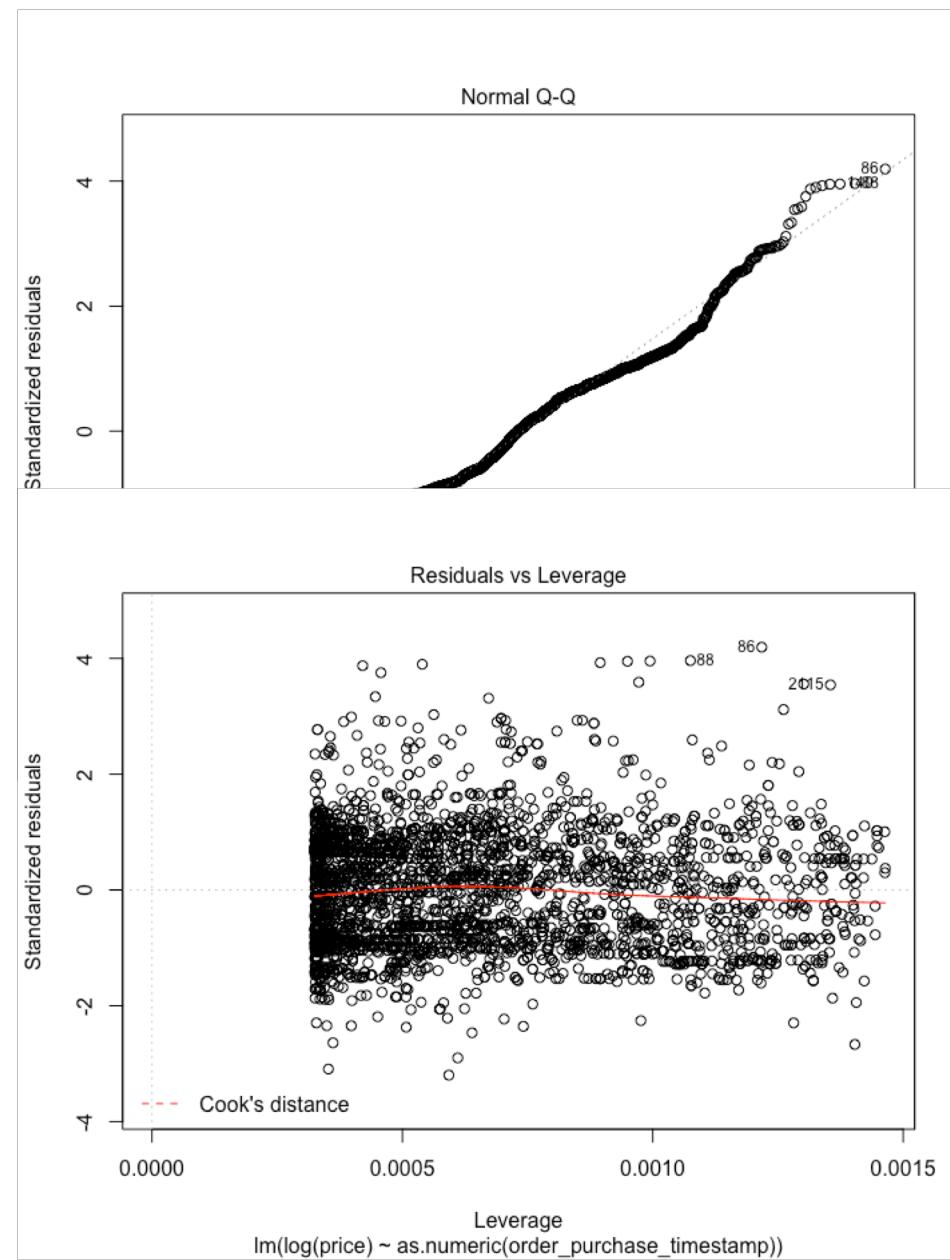
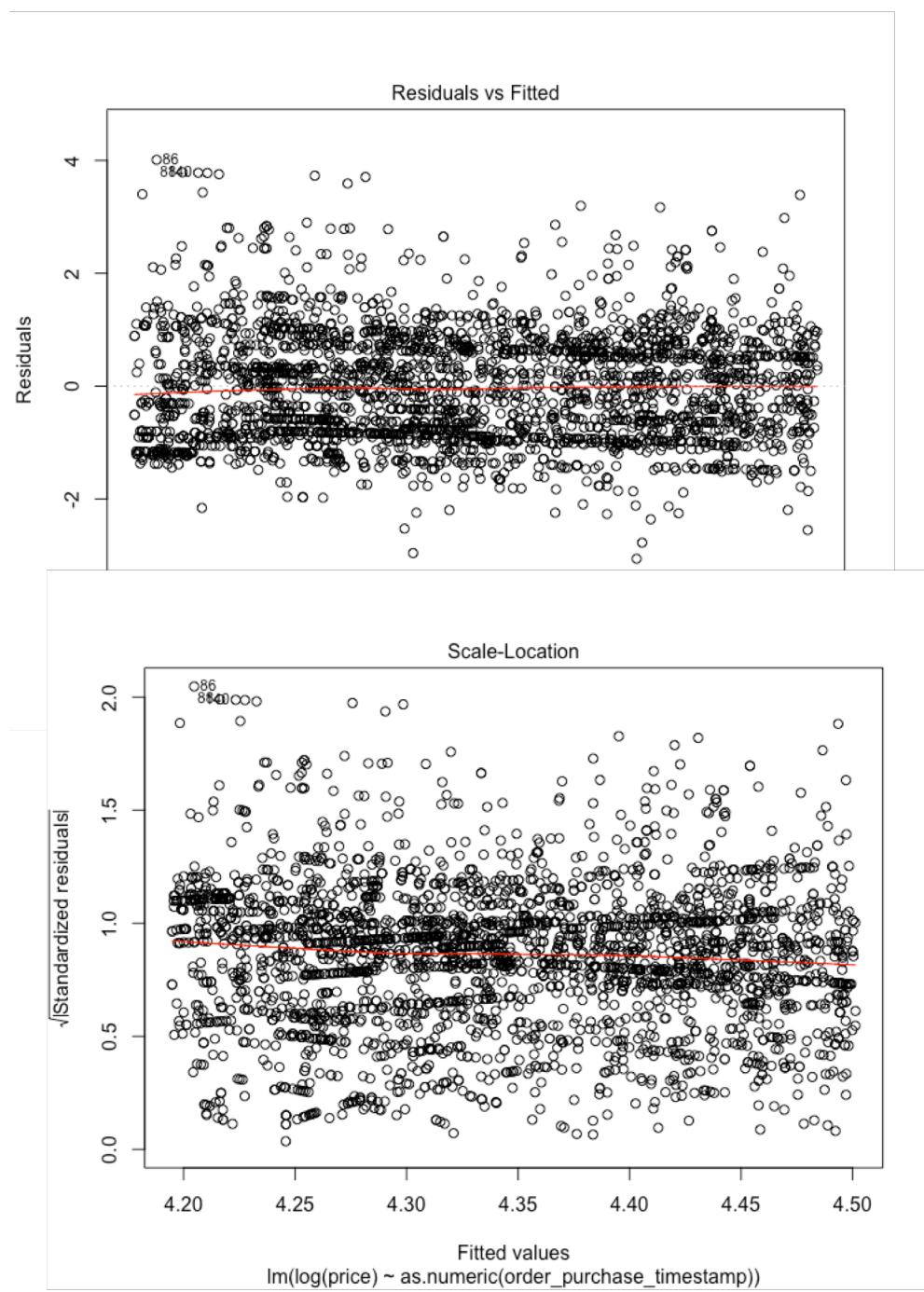
```
Call:  
lm(formula = (price) ~ as.numeric(order_purchase_timestamp),  
  data = main_df_comp_ass)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-128.2   -96.7   -56.4    23.5  3564.1  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 1.363e+02 8.513e+00 16.008 <2e-16 ***  
as.numeric(order_purchase_timestamp) -2.269e-04 3.363e-04 -0.675     0.5  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 228.6 on 3070 degrees of freedom  
Multiple R-squared:  0.0001482, Adjusted R-squared:  -0.0001775  
F-statistic: 0.4551 on 1 and 3070 DF,  p-value: 0.5
```

Line



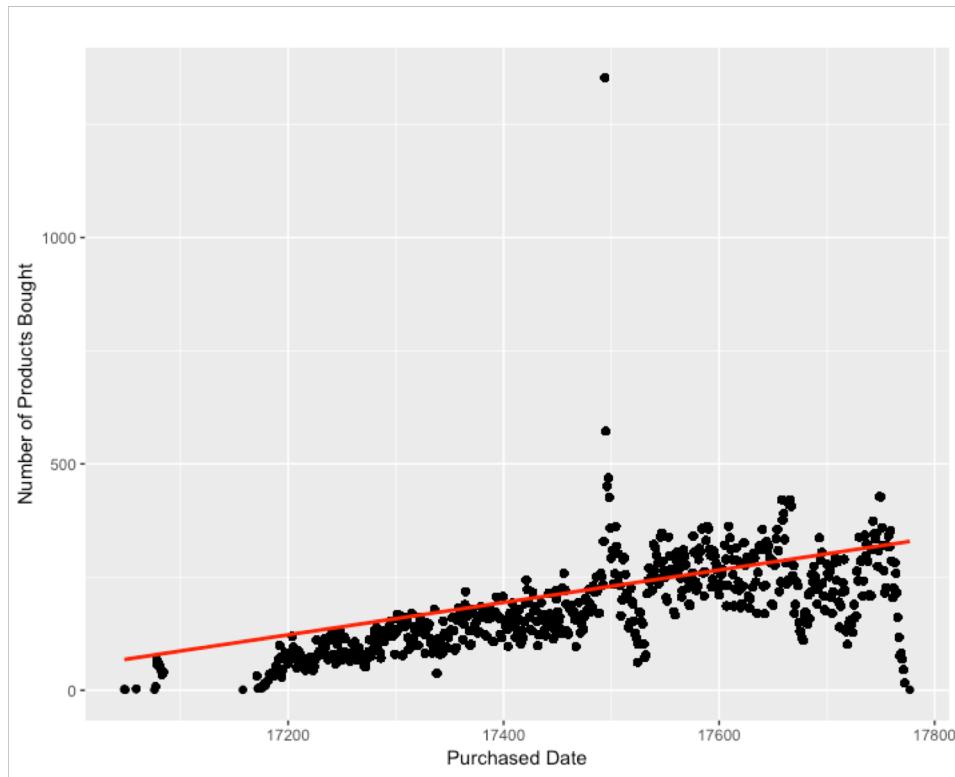
Log Regression





```
Call:  
lm(formula = log(price) ~ as.numeric(order_purchase_timestamp),  
  data = main_df_comp_ass)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-3.0592 -0.8034  0.0088  0.7021  4.0114  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 4.192e+00 3.566e-02 117.57 < 2e-16 ***  
as.numeric(order_purchase_timestamp) 6.859e-06 1.408e-06   4.87 1.17e-06 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 0.9573 on 3070 degrees of freedom  
Multiple R-squared:  0.007666, Adjusted R-squared:  0.007343  
F-statistic: 23.72 on 1 and 3070 DF, p-value: 1.172e-06
```

Purchased Date to Frequency



Prospective Usage

- Prospecting small businesses.
- Businesses can provide promotions on the least frequent products in their respective areas to boost sales.