Name: SAURABH SHEKHAR PISAL   Contact: +1-312-804-7225   Email: spisal@hawk.iit.edu

CAR DASH ASSESSMENT

Report By :   Saurabh Shekhar Pisal
Contact Details: +1-312-804-7225
Email Id: spisal@hawk.iit.edu

LinkedIn Profile: https://www.linkedin.com/in/saurabh-shekhar-pisal-7a370637/   Tableau Profile: https://public.tableau.com/profile/saurabh.shekhar.pisal

Name: SAURABH SHEKHAR PISAL   Contact: +1-312-804-7225   Email: spisal@hawk.iit.edu

Contents

Part 1

Name: *SAURABH SHEKHAR PISAL*   Contact: *+1-312-804-7225*   Email: *spisal@hawk.iit.edu*

1. Provide a set of recommendations on how to improve our business or product based on the attached dataset. This is intended to be fairly open-ended - there's no right or wrong answer! We're more concerned with your approach and the insights you uncover.

CarDash Revenue sheet
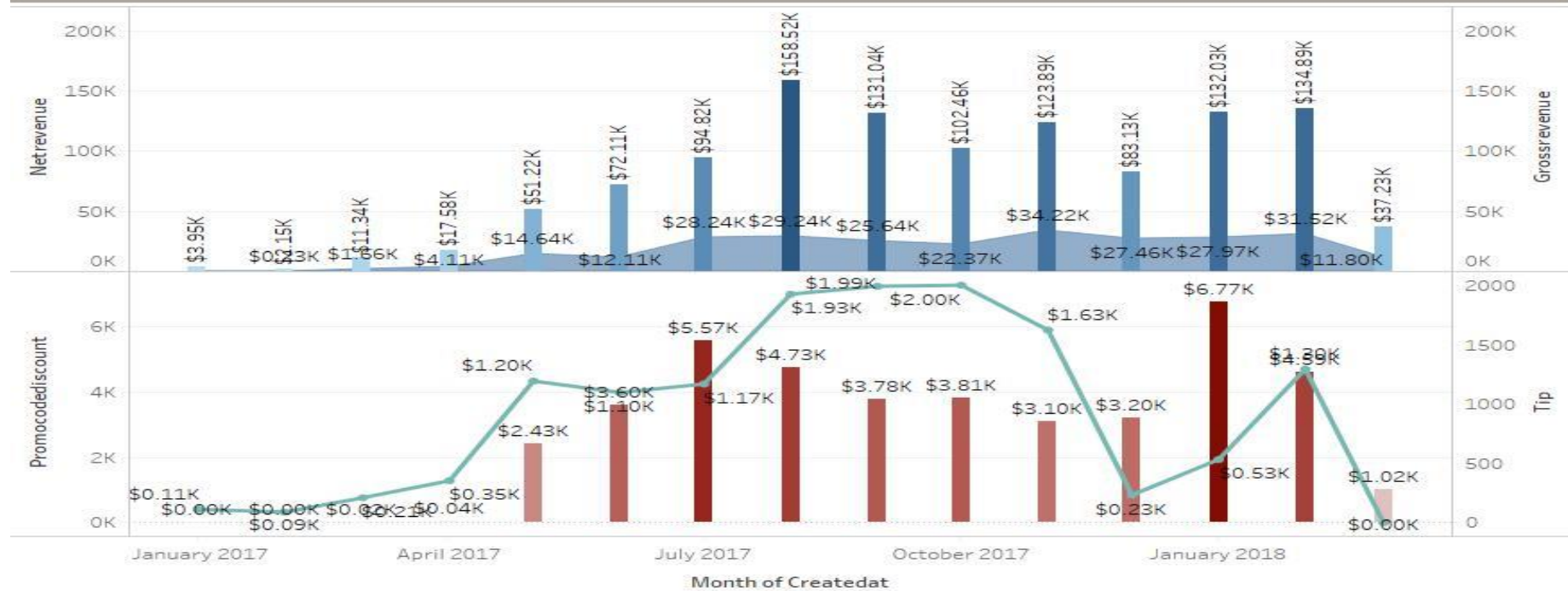
The trends of Netrevenue, sum of Grossrevenue, sum of Promocodediscount and Tip for Createdat Month. For pane Sum of Grossrevenue: Color shows sum of Grossrevenue. The marks are labeled by sum of Grossrevenue. For pane Sum of Tip: Color shows details about Netrevenue and Tip. The marks are labeled by Tip. For pane Sum of Promocodediscount: Color shows sum of Promocodediscount. The marks are labeled by sum of Promocodediscount. For pane Sum of Netrevenue: Color shows details about Netrevenue and Tip. The marks are labeled by Netrevenue. The data is filtered on Createdat Month, which keeps 12 of 12 members.

Measure Names
■ Netrevenue
■ Tip

Promocodediscount
0   6,771

Grossrevenue
2,154   158,516

Name: SAURABH SHEKHAR PISAL   **Contact: +1-312-804-7225**   **Email: spisal@hawk.iit.edu**

By taking into consideration the data provided by Car Dash, I have made the Above Visual analysis (CarDash Revenue Sheet) .

Net-revenue VS Gross-Revenue

From the initial analysis we can find out that the net revenue of the company goes hand in hand with the gross revenue. I have clubbed the data into Monthly data (Taking into consideration each year). But this analysis is not going to help the company make any prediction of the future. This Visual analysis only states that the company does well and earns good net revenue if the Gross revenues are high.

Promotion-Discount vs Tip

As we can see from the visualization, the TIP coming from the customer is likely to come more if the Promotion-Discount given to the customer is more. Hence, investing in Promotion-Discounts – A strategy for customer retrieval and growing the bucket of customers coming back any marketing campaign like promotion discounts will help the company get new as well as old customers. This factor of Promotion-Discounts is helping the company grow in long term.

Name: SAURABH SHEKHAR PISAL   Contact: +1-312-804-7225   Email: spisal@hawk.iit.edu

Promotion affecting the NetRevenue
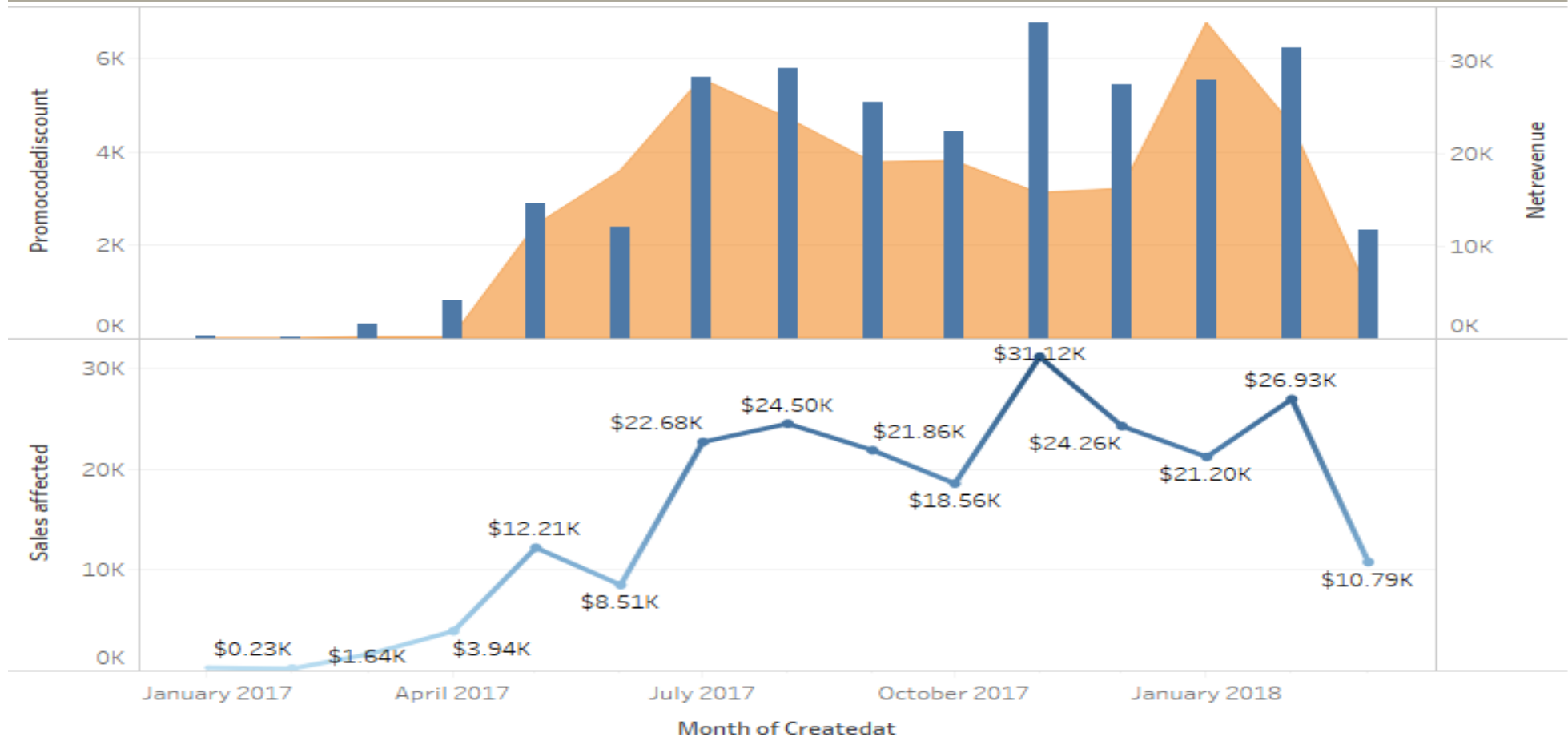
The trends of Promocodediscount, Netrevenue and sum of Sales affected for Createdat Month. For pane Sum of Promocodediscount: Color shows details about Promocodediscount and Netrevenue. For pane Sum of Netrevenue: Color shows details about Promocodediscount and Netrevenue. For pane Sum of Sales affected: Color shows sum of Sales affected. The marks are labeled by sum of Sales affected. The data is filtered on Createdat Month, which keeps 12 of 12 members.

Measure Names
- ■ Netrevenue
- ■ Promocodediscount

Sales affected

231   31,118

Name: SAURABH SHEKHAR PISAL   **Contact: +1-312-804-7225**   **Email: spisal@hawk.iit.edu**

Promocodediscount vs Netrevenue

With respect to the above figure (Promotion Affecting the NetRevenue) I tried to check the effect of PromotioncodeDiscount on net revenue and it gives a positive result. There is grown in the net revenue every time when there is a good Promocodediscount given to the customers in that particular month. If there is less PromotioncodeDiscount in any month for example between July 2017 – October 2017 we can see a drop in NetRevenue with Drop in PromotionCodeDiscounts. Here we find a Linear relationship between the Promocodediscount and NetRevenue of the company.

Sales Affected each Month of the year

To test the same, I created a new Column (A calculated field where the difference between netrevenue and the amount of discount was calculated) to check how it is affecting the sales for that particular month of the year. If we compare here the Sales affected line graph and the PromotioncodeDiscount Vs Netrevenue it can be clearly seen that PromotioncodeDiscount plays a major role in affecting the netrevenue of the company.

Name: SAURABH SHEKHAR PISAL   Contact: +1-312-804-7225   Email: spisal@hawk.iit.edu

2.Choose one of the recommendations/insights you uncovered (in #1) and outline an experiment you would like to run to test your suggested product/business recommendation. Please state your hypothesis, describe how you would structure your experiment, list your success metrics and describe the implementation.

To prove my hypothesis, I will be building a linear regression model to test the effect of the independent variables (closingdate, tip, promocodediscount)

I will not be taking into consideration the Grossrevenue, as it is a dependent variable and net revenue is derived from Gross revenue.

I will not be taking into consideration the FinalInvoice as it is a part of the net revenue and has direct relation ship with it.

I will be considering those independent variables which affect the net revenue directly as an input and a model can be predicted from it to Finalize my hypothesis derived from #1 Question.

For the same I will be building a model In R code and test its various parameters.(Taking into consideration Confidence level = 95%)

Analysis 1

```
> car_dash_model<-lm(netrevenue~as.factor(closingdate)+as.factor(pickupdate)+tip+promocodediscount)
> summary(car_dash_model)

Call:
lm(formula = netrevenue ~ as.factor(closingdate) + as.factor(pickupdate) +
    tip + promocodediscount)

Residuals:
    Min      1Q  Median      3Q     Max
-1405.6   -26.3     1.9     1.9  3586.4

Coefficients: (17 not defined because of singularities)
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                            -1.852e+00  2.955e+00  -0.627 0.530803
as.factor(closingdate)2017-08-17 00:00:00 -2.209e+02  8.527e+01  -2.590 0.009613 **
as.factor(closingdate)2017-08-18 00:00:00 -5.612e+01  6.914e+01  -0.812 0.417023
as.factor(closingdate)2017-08-19 00:00:00  8.412e+01  9.829e+01   0.856 0.392119
as.factor(closingdate)2017-08-20 00:00:00  9.280e+00  8.390e+01   0.111 0.911936
as.factor(closingdate)2017-08-21 00:00:00  1.801e+02  6.994e+01   2.575 0.010061 *
as.factor(closingdate)2017-08-22 00:00:00  2.752e+02  1.078e+02   2.554 0.010674 *
as.factor(closingdate)2017-08-23 00:00:00 -1.242e+01  7.948e+01  -0.156 0.875842
as.factor(closingdate)2017-08-24 00:00:00 -1.200e+01  8.740e+01  -0.137 0.890759
as.factor(closingdate)2017-08-25 00:00:00  9.589e+01  9.634e+01   0.995 0.319606
```

Name: SAURABH SHEKHAR PISAL   Contact: +1-312-804-7225   Email: spisal@hawk.iit.edu

```
as.factor(pickupdate)2018-04-06 00:00:00    1.852e+00   1.390e+02    0.013 0.989365
as.factor(pickupdate)2018-04-16 00:00:00   -8.864e+00   1.390e+02   -0.064 0.949142
as.factor(pickupdate)2018-04-17 00:00:00   -8.864e+00   1.390e+02   -0.064 0.949142
as.factor(pickupdate)2018-04-18 00:00:00    1.852e+00   1.390e+02    0.013 0.989365
as.factor(pickupdate)2018-05-02 00:00:00    1.852e+00   1.390e+02    0.013 0.989365
as.factor(pickupdate)2018-12-10 00:00:00   -1.958e+01   1.390e+02   -0.141 0.887971
tip                                          2.272e+00   2.776e-01    8.185 3.29e-16 ***
promocodediscount                            1.072e+00   1.714e-01    6.253 4.30e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 138.9 on 6078 degrees of freedom
Multiple R-squared:  0.4299,    Adjusted R-squared:  0.3846
F-statistic: 9.488 on 483 and 6078 DF,  p-value: < 2.2e-16

> |
```

From the above analysis we realize that the Pickupdate doesn't affect the model as the values of its Beta parameter is more than 0.05(CL=95%) and Tip and Promocodediscount have value less than 0.05.

Here the F statistics value to be noted is 9.488 and adjust-Rsquare = 0.3846

Now, we will again build a model by eliminating the independent variables not affecting the dependent variable (Netrevenue)

Analysis 2

By Eliminating the pickup date and closing date we get the below analysis

```
> car_dash_model<-lm(netrevenue~tip+promocodediscount)
> summary(Car_dash_model)
Error in summary(Car_dash_model) : object 'Car_dash_model' not found
> summary(car_dash_model)

Call:
lm(formula = netrevenue ~ tip + promocodediscount)

Residuals:
    Min      1Q  Median      3Q     Max
-1797.1   -27.6   -27.6     1.4  4260.8

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        27.6407     2.5309  10.921  < 2e-16 ***
tip                 2.6696     0.2789   9.574  < 2e-16 ***
promocodediscount   1.2470     0.1924   6.480 9.82e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 175.1 on 6559 degrees of freedom
Multiple R-squared:  0.02251,   Adjusted R-squared:  0.02221
F-statistic: 75.52 on 2 and 6559 DF,  p-value: < 2.2e-16

> |
```

Here we can not that the value of tip and promocodediscount is less than 0.05(Taking into consideration CL=95%).

Hence, we can see that tip and promocode discount affect the netrevenue of the company directly.

F-Statistic value = 75.52 which is very much greater than previous model(PART 1). As well as the oval all P-value of the model is less than 0.05. Hence this is the best fitted prediction model for the netrevenue of the company which can be used by the company for decision making.
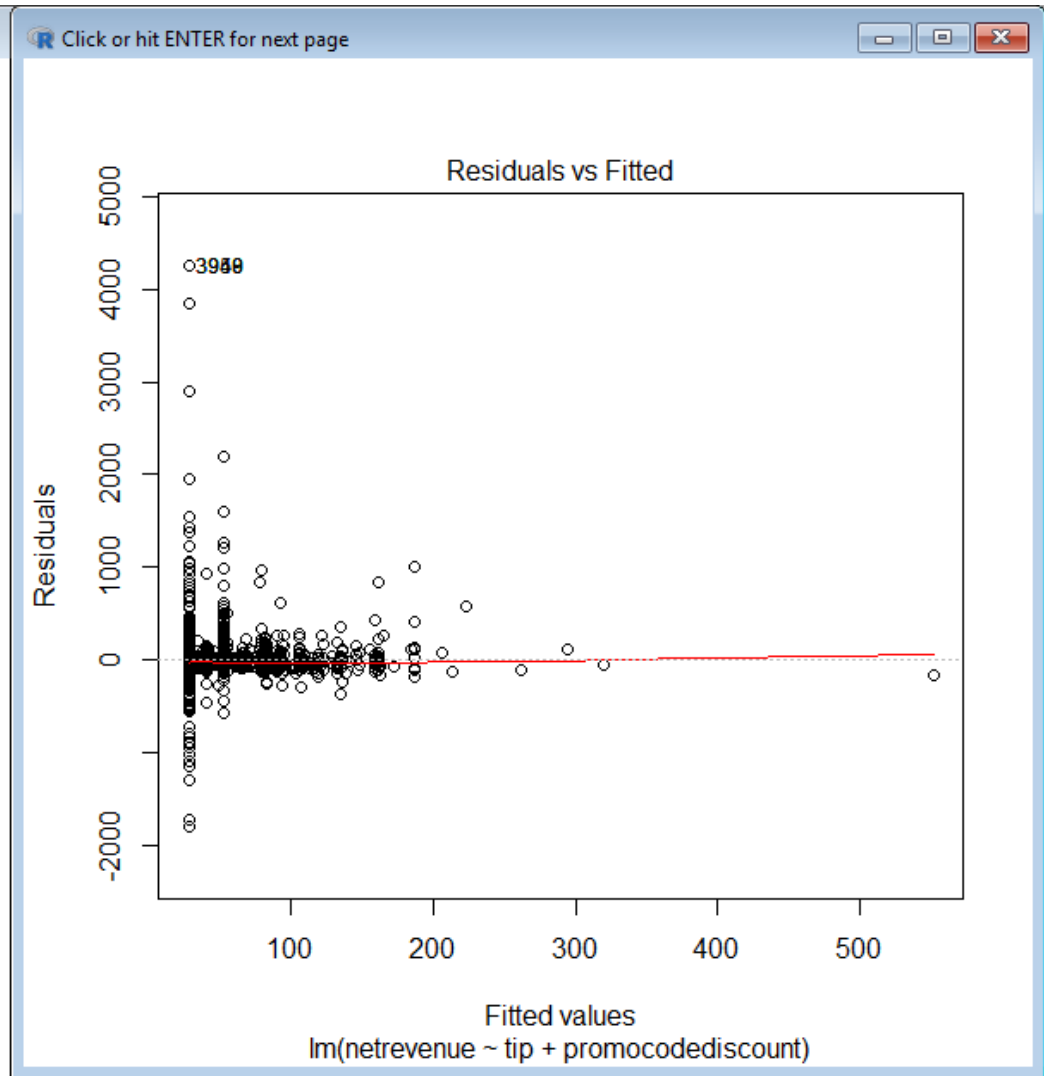
The fitted model is as below

Netrevenue = 27.6407 + 2.6696 * [TIP] + 1.2470 * [PROMOCODEDISCOUNT] + e

Name: *SAURABH SHEKHAR PISAL*   Contact: *+1-312-804-7225*   Email: *spisal@hawk.iit.edu*

To further test the linearity of the fitted regression model we do further testing.

Residual Analysis of the model

From the plot of residual vs, the X variables (tip ,promocodediscount) we can see that the red line is fairly linear, and we can conclude that there is a constant variance . This test proves that the model is the best model.

Name: *SAURABH SHEKHAR PISAL*   Contact: *+1-312-804-7225*   Email: *spisal@hawk.iit.edu*
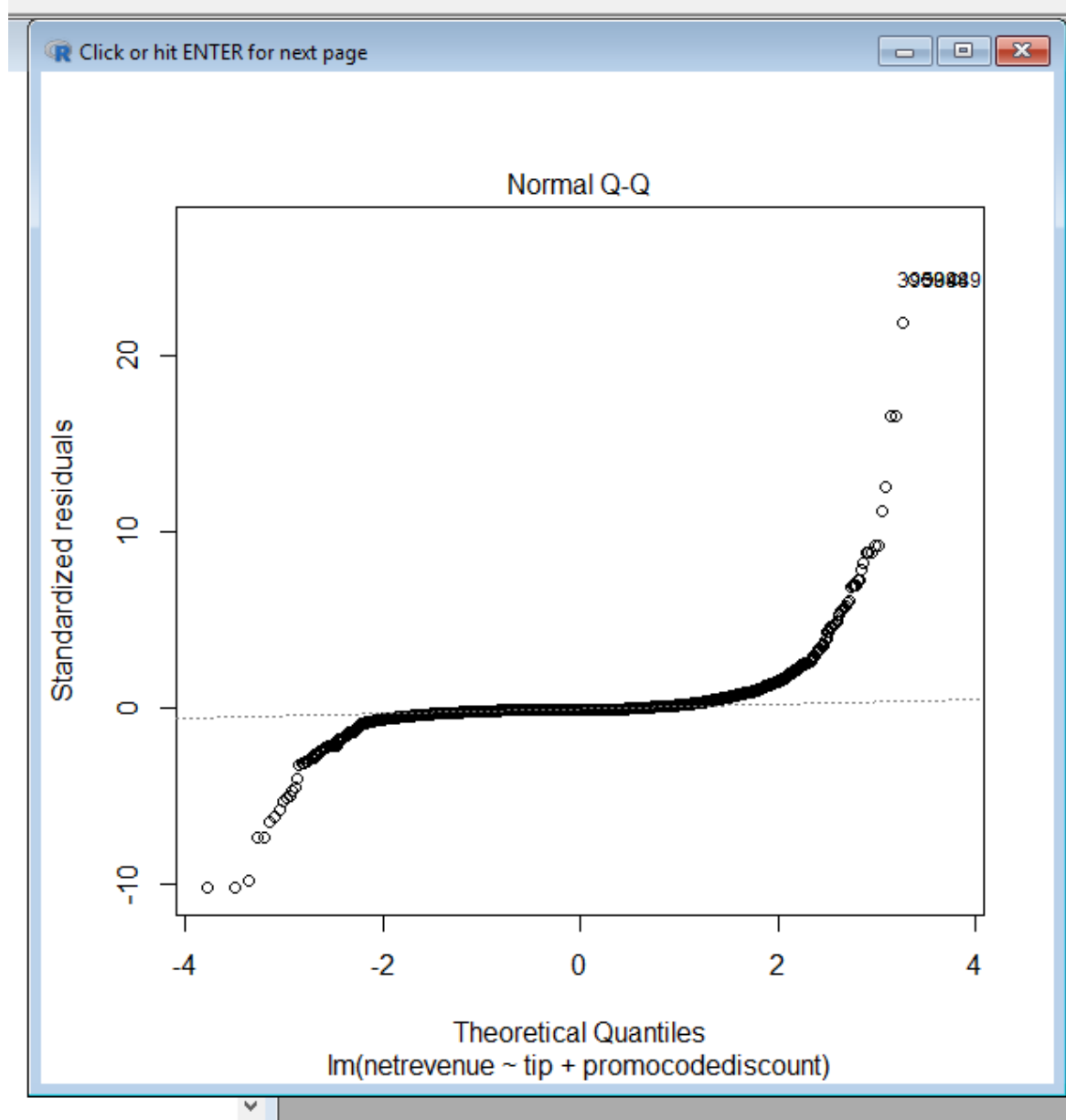
Normal QQ plot

From the normal QQ plot we can see that the almost all the variables lie on the line plotted. Hence, the fitted regression model follows Normal Distribution

Name: SAURABH SHEKHAR PISAL   Contact: +1-312-804-7225   Email: spisal@hawk.iit.edu

3. Let's assume that the experiment you ran (in #2) proved your hypothesis was true. How would you suggest implementing the change on a larger scale? What are some operational challenges you might encounter and how would you mitigate their risk?

According to my Hypothesis and Analysis from #2 we all can conclude that the net revenue of Car Dash is highly affected by the tips coming in from the customers per month and the Promocodediscount we give per month. Hence, to increase the revenue more of the company we need to focus on customer retention methodologies. For example we need to have more marketing of the company's products and its services and along with that in order to attract more customers and retain the existing customer we need to give them schemes and tactics which will make them spend money and have marketing skills within the organization to offer them for more in case a customer comes back. For example, if we gave a promocodediscount on the additional car speakers and lure them into buying another product and giving a discount to them on the other product if they buy the car speakers.

In a net sum, the company is selling more, and the customers are coming back, and even more customers are adding into the customers bucket.

To mitigate the risk of customers not returning we will have to keep schemes in such a way that we stay in the profit margin zone and the customers keep coming back and keep spending due to the offers. Another way is to keep and expiration on the promocode in our Profit cycle (As per requirement) so that we know where we can stay during a Quarterly turn over(Predict) and for the same, the prediction model which I have designed above can be used.

Part 2

Name: SAURABH SHEKHAR PISAL   **Contact: +1-312-804-7225**   **Email: spisal@hawk.iit.edu**

Assumptions & Facts

-- table Consumer has Consumer_ID

-- assumption of new column is Join_Date: When the consumer first registered

-- TABLE Order

--Has ID, Creator_ID==Consumer_ID,

--assumption of new column is Order_Date: When was the order placed by the customer

--A. QUERY TO ESTABLISH MONTHLY COHORTS BASED ON MONTH JOINED

select con.Join_Date as Join_Date,count(DISTINCT ord.ID) as Order_counts from Order as ord

join Consumer as con on ord.Creator_ID=con.Consumer_ID

group by Join_Date

order by Join_date;

--B. queries to calculate weighted average retention of new customers by months after first purchase

--first, we will find the difference in their activity, and order it in ascending order which will show average retention of new customers by months after their first order

select PERIOD_DIFF(DATE_FORMAT(ord.Order_date,'%Y%m'),DATE_FORMAT(con.Join_Date,'%Y%m')) as Month_Number,con.Consumer_ID

from Order as ord

join Consumer as con on ord.Creator_ID=con.Consumer_ID

group by Month_Number,con.Consumer_ID

order by Month_Number;

Name: SAURABH SHEKHAR PISAL   **Contact: +1-312-804-7225**   **Email: spisal@hawk.iit.edu**

Interactive Charts on Tableau Profile Link

https://public.tableau.com/profile/saurabh.shekhar.pisal#!/vizhome/CarDashfactorsaffectingitsRevenue/Dashboard1?publish=yes