Presented By:
Bharati Malik

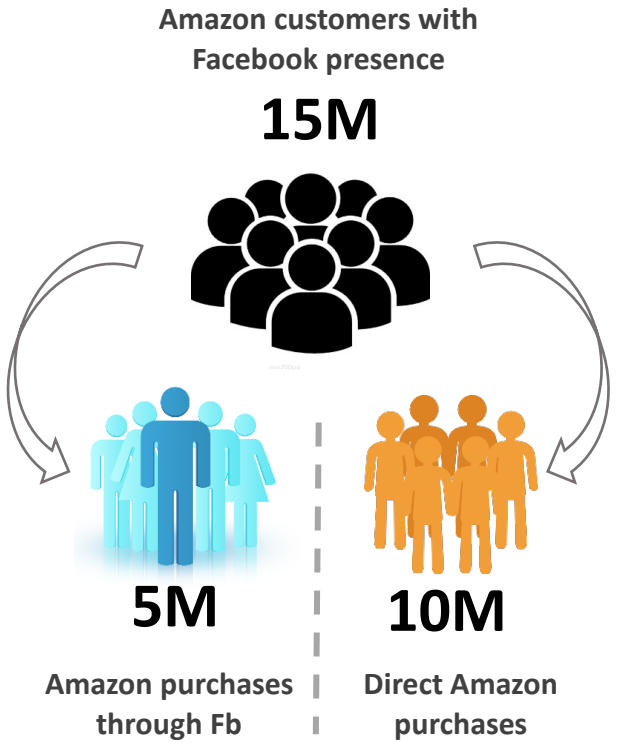
Agenda

- Business Context
- Research Question
- Data Overview
- Analysis
- Insights
- Recommendations
- Limitations

Business Context

Amazon relies heavily on digital marketing:

- DYNAMIC – Reach and Engage
- TRACKABLE – What works and what doesn't, improvise
- AFFORDABLE – Higher ROI compared to traditional marketing



Facebook gets the biggest share with an annual spend of ***$100M***!!

Research Question

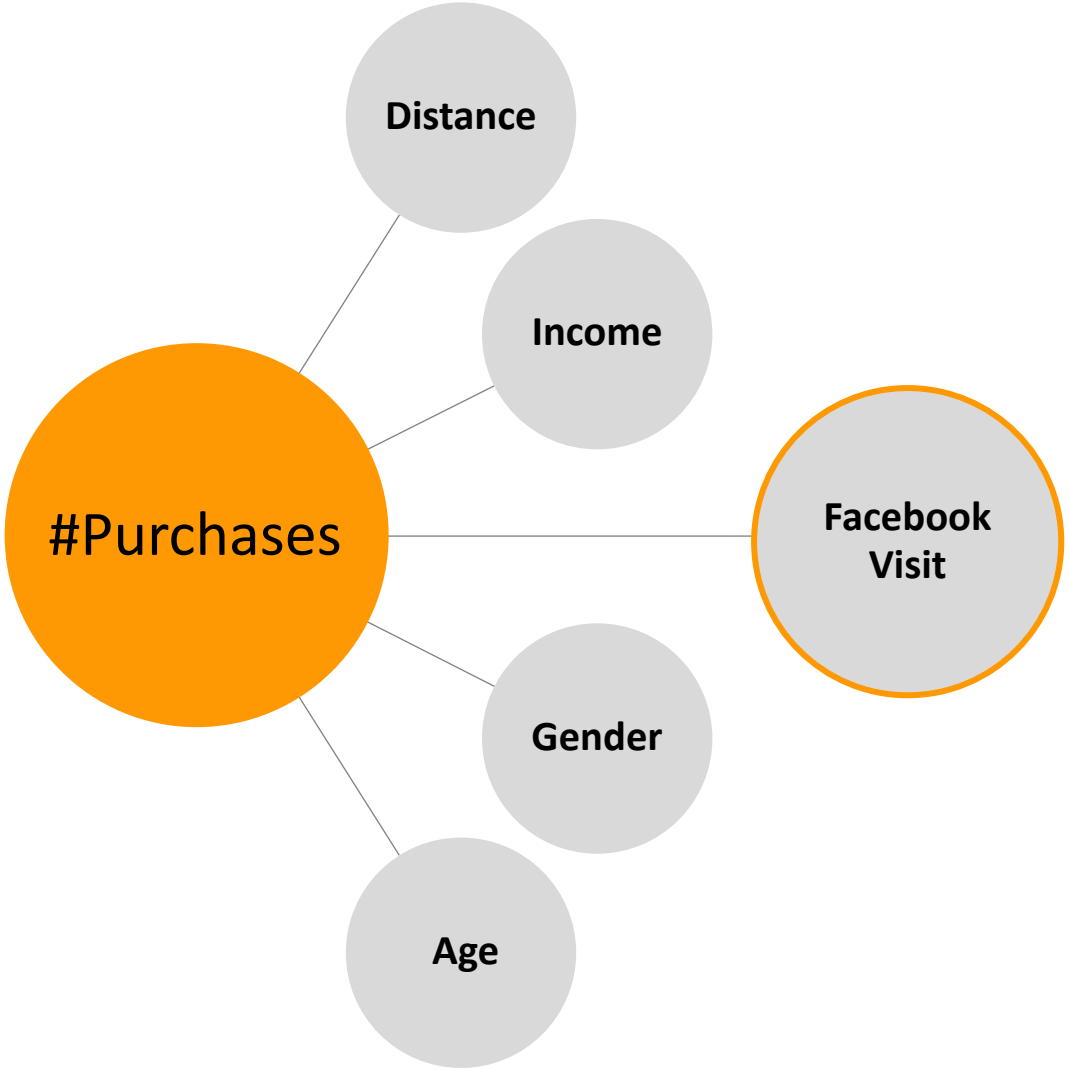
What is the impact of Facebook advertising on Amazon's revenue?

Is it ***4X*** the advertising cost?

100M

400M

Data Overview



Facebook visit

All the purchases after Facebook visit

All the purchases directly through Amazon

1225 **2422**

Sample data available for customers from **Northern California (3677)**

Analysis

Descriptive statistics by group

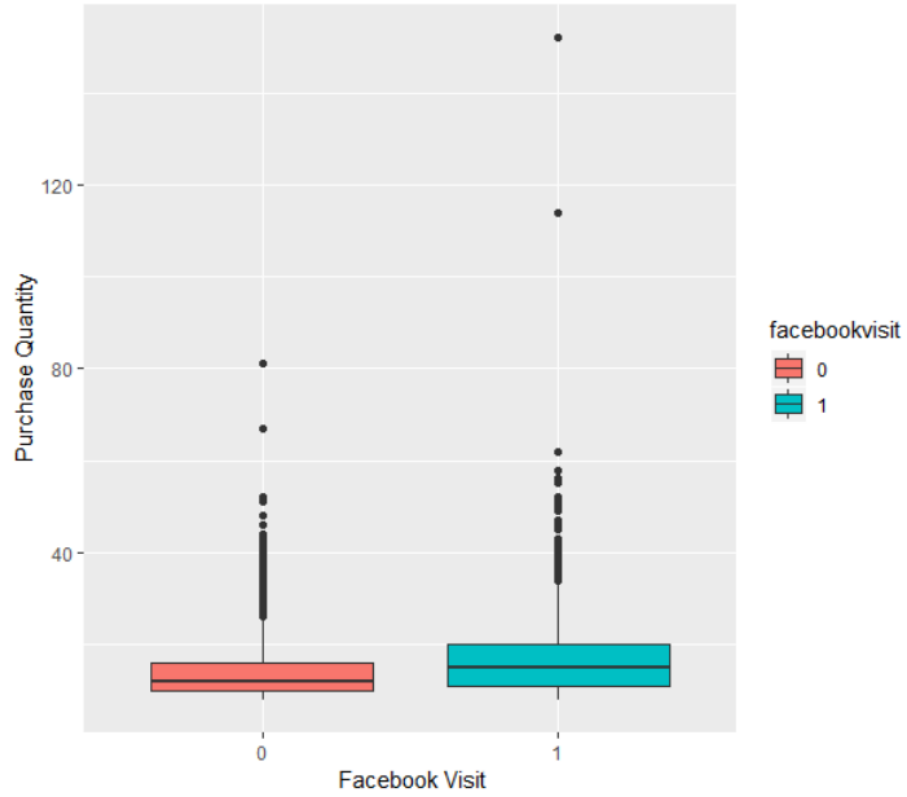
group: 0 (Purchase directly through Amazon)

	vars	n	mean	sd	median	trimmed	mad	min	max	range
distance	1	2452	1.50	1.18	1	1.41	1.48	0	6	6
purchase	2	2452	13.83	6.09	12	12.87	4.45	8	81	73
income	3	2452	11.69	3.69	12	12.05	2.97	0	17	17
gender	4	2452	0.10	0.31	0	0.01	0.00	0	1	1
facebookvisit	5	2452	0.00	0.00	0	0.00	0.00	0	0	0
cust_age	6	2452	33.25	5.99	32	32.87	7.41	25	50	25

---

group: 1 (Purchase after Facebook visit)

	vars	n	mean	sd	median	trimmed	mad	min	max	range
distance	1	1225	2.54	1.40	2	2.48	1.48	0	8	8
purchase	2	1225	16.80	9.17	15	15.46	5.93	8	152	144
income	3	1225	10.16	3.90	11	10.37	2.97	0	17	17
gender	4	1225	0.29	0.45	0	0.24	0.00	0	1	1
facebookvisit	5	1225	1.00	0.00	1	1.00	0.00	1	1	0
cust_age	6	1225	36.24	6.65	36	36.43	8.90	25	50	25



Analysis

	Dependent variable:
	purchase
	Negetive Binomial Results
distance	1.1327***
	(0.0059)
income	1.0147***
	(0.0021)
cust_age	1.0010
	(0.0012)
facebookvisit	1.0816***
	(0.0188)
gender	1.0176
	(0.0256)
Constant	9.2274***
	(0.0468)
Observations	3,677
Log Likelihood	-11,331.9500
theta	12.2743*** (0.4867)
Akaike Inf. Crit.	22,675.9000
Note:	*p<0.05; **p<0.01; ***p<0.001

Number of purchases is ***8.16%*** more when the purchase is made though Facebook vs when it is made directly from Amazon.

```
Likelihood ratio test

Model 1: purchase ~ 1
Model 2: purchase ~ distance + income + cust_age + facebookvisit + gender
  #Df LogLik Df  Chisq Pr(>Chisq)
1   2 -11742
2   7 -11331  5 821.55  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Insights

204M additional revenue from purchases made through Facebook

Recommendations

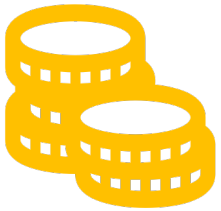
- Renegotiate with Facebook to reduce Ads spending to $50M or less.

Advertising Cost of Spend (ACoS)	=	Total Ad Spend/ Total Ad Sales	=	100/400	=	$0.25

- Collect and analyze more data to better quantify impact of Facebook Ads on revenue. E.g. user might see an Ad on Facebook and later buy directly from Amazon.

Limitations

Data is limited to **Northern California Region.**



Transaction Amount for every customer is unknown.

No data on **Ad Frequency** that led to a direct Amazon purchase post **Facebook Ad Exposure.**

Thank you!

Appendix

Final Model

$$\text{Purchase} = \beta 0 + \beta 1\ \text{facebook visit} + \beta 2\ \text{income} + \beta 3\text{gender} + \beta 4\text{distance} + \beta 5\text{cust\_age}$$

Summary Statistics – All variables

```
 Descriptive statistics by group
group: 0
               vars    n   mean    sd median trimmed   mad min max range
income            1 2452  11.69  3.69     12   12.05  2.97   0  17    17
gender            2 2452   0.10  0.31      0    0.01  0.00   0   1     1
facebookvisit     3 2452   0.00  0.00      0    0.00  0.00   0   0     0
distance          4 2452   1.50  1.18      1    1.41  1.48   0   6     6
cust_age          5 2452  33.25  5.99     32   32.87  7.41  25  50    25
purchase          6 2452  13.83  6.09     12   12.87  4.45   8  81    73
numoffriends      7 2452  99.07 19.84     99   99.09 19.27  31 165   134
numofposts        8 2452 199.65 34.39    199  199.44 34.10  91 309   218
PublicProfile     9 2452   0.27  0.45      0    0.22  0.00   0   1     1
------------------------------------------------------------------------
group: 1
               vars    n   mean    sd median trimmed   mad min max range
income            1 1225  10.16  3.90     11   10.37  2.97   0  17    17
gender            2 1225   0.29  0.45      0    0.24  0.00   0   1     1
facebookvisit     3 1225   1.00  0.00      1    1.00  0.00   1   1     0
distance          4 1225   2.54  1.40      2    2.48  1.48   0   8     8
cust_age          5 1225  36.24  6.65     36   36.43  8.90  25  50    25
purchase          6 1225  16.80  9.17     15   15.46  5.93   8 152   144
numoffriends      7 1225 129.19 19.62    129  128.97 19.27  74 206   132
numofposts        8 1225 215.24 35.91    215  215.43 35.58 100 344   244
PublicProfile     9 1225   0.31  0.46      0    0.26  0.00   0   1     1
```

Multicollinearity - Model Variables

	distance	income	gender	facebookvisit	cust_age
distance	1.000	-0.068	0.148	0.364	0.098
income	-0.068	1.000	-0.348	-0.188	-0.098
gender	0.148	-0.348	1.000	0.235	0.026
facebookvisit	0.364	-0.188	0.235	1.000	0.221
cust_age	0.098	-0.098	0.026	0.221	1.000

```
No variable from the 5 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( cust_age ~ gender ):  0.02563077
max correlation ( facebookvisit ~ distance ):  0.3641455

---------- VIFs of the remained variables --------
      Variables      VIF
1      distance 1.159531
2        income 1.159732
3        gender 1.186635
4 facebookvisit 1.263052
5      cust_age 1.058062
```

Endogeneity

Omitted Variable: Technology Friendliness
Endogenous Variable: Facebook Visit

Instrument Variables:

1. Number of posts
2. Number of friends
3. Public Profile

```
ivreg(formula = purchase ~ facebookvisit + distance + income +
    cust_age + gender | numoffriends + numofposts + PublicProfile +
    distance + income + cust_age + gender, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max
-15.330  -3.829  -1.456   2.161 135.587

Coefficients:
             Estimate Std. Error t value         Pr(>|t|)
(Intercept)  8.186421   0.761217  10.754          < 2e-16 ***
facebookvisit 1.407317   0.475982   2.957          0.00313 **
distance     1.895852   0.099650  19.025          < 2e-16 ***
income       0.219083   0.031933   6.861 0.00000000000801 ***
cust_age     0.005273   0.018824   0.280          0.77939
gender       0.256071   0.336440   0.761          0.44663

Diagnostic tests:
                 df1  df2 statistic p-value
Weak instruments   3 3669   565.519  <2e-16 ***
Wu-Hausman         1 3670     0.208   0.649
Sargan             2   NA     2.225   0.329
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.807 on 3671 degrees of freedom
Multiple R-Squared: 0.1539,     Adjusted R-squared: 0.1528
Wald test: 131.2 on 5 and 3671 DF,  p-value: < 2.2e-16
```

Endogeneity – IV correlation check

	numoffriends	numofposts	PublicProfile	purchase
numoffriends	1.000	0.107	0.006	0.105
numofposts	0.107	1.000	0.012	0.061
PublicProfile	0.006	0.012	1.000	0.025
purchase	0.105	0.061	0.025	1.000

```
No variable from the 4 input variables has collinearity problem.

The linear correlation coefficients ranges between:
min correlation ( PublicProfile ~ numoffriends ):  0.006197103
max correlation ( numofposts ~ numoffriends ):  0.1065741

---------- VIFs of the remained variables --------
      Variables      VIF
1  numoffriends 1.021464
2    numofposts 1.014214
3 PublicProfile 1.000753
4      purchase 1.014286
```

Negative Binomial vs Poisson

```
> lrtest(poisson1, poissonempty) # possion is not a good fit
Likelihood ratio test

Model 1: purchase ~ distance + income + cust_age + facebookvisit + gender
Model 2: purchase ~ 1
  #Df LogLik Df  Chisq Pr(>Chisq)
1   6 -12416
2   1 -13428 -5 2024.4  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Comparison with negative binomial

```
Likelihood ratio test

Model 1: purchase ~ distance + income + cust_age + facebookvisit + gender
Model 2: purchase ~ distance + income + cust_age + facebookvisit + gender
  #Df LogLik Df  Chisq Pr(>Chisq)
1   6 -12416
2   7 -11331  1 2169.1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

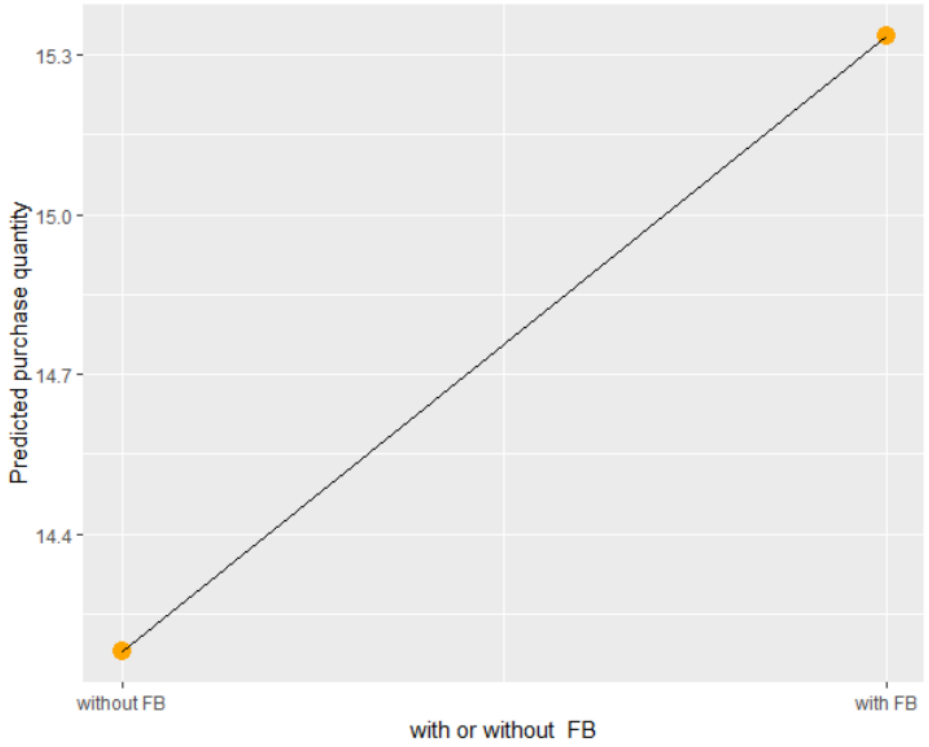
Negative Binomial vs Ideal model

```
Likelihood ratio test

Model 1: purchase ~ 1
Model 2: purchase ~ distance + income + cust_age + facebookvisit + gender
  #Df LogLik Df  Chisq Pr(>Chisq)
1   2 -11742
2   7 -11331  5 821.55  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Heteroskedasticity

```
> gqtest(negbin1) # Significant Goldfeld-Quandt test does not indicate heteroscedasticity

        Goldfeld-Quandt test

data:  negbin1
GQ = 0.94104, df1 = 1833, df2 = 1832, p-value = 0.9033
alternative hypothesis: variance increases from segment 1 to 2

> bptest(negbin1) # Significant Breusch-Pagan test  indicates heteroscedasticity

        studentized Breusch-Pagan test

data:  negbin1
BP = 30.567, df = 5, p-value = 0.0000114
```

Predicted values



	facebookvisit	cust_age	distance	gender	income	predicted_purchase
1	0	34.24476	1.843351	0.166712	11.18031	14.17957
2	1	34.24476	1.843351	0.166712	11.18031	15.33611

Extras

Descriptive Statistics - combined

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
income	3,677	11.2	3.8	0	9	12	13	17
gender	3,677	0.2	0.4	0	0	0	0	1
facebookvisit	3,677	0.3	0.5	0	0	0	1	1
distance	3,677	1.8	1.4	0	1	2	3	8
cust_age	3,677	34.2	6.4	25	29	34	39	50
purchase	3,677	14.8	7.4	8	10	13	17	152
numoffriends	3,677	109.1	24.3	31	92	108	125	206
numofposts	3,677	204.8	35.7	91	181	204	229	344
PublicProfile	3,677	0.3	0.5	0	0	0	1	1

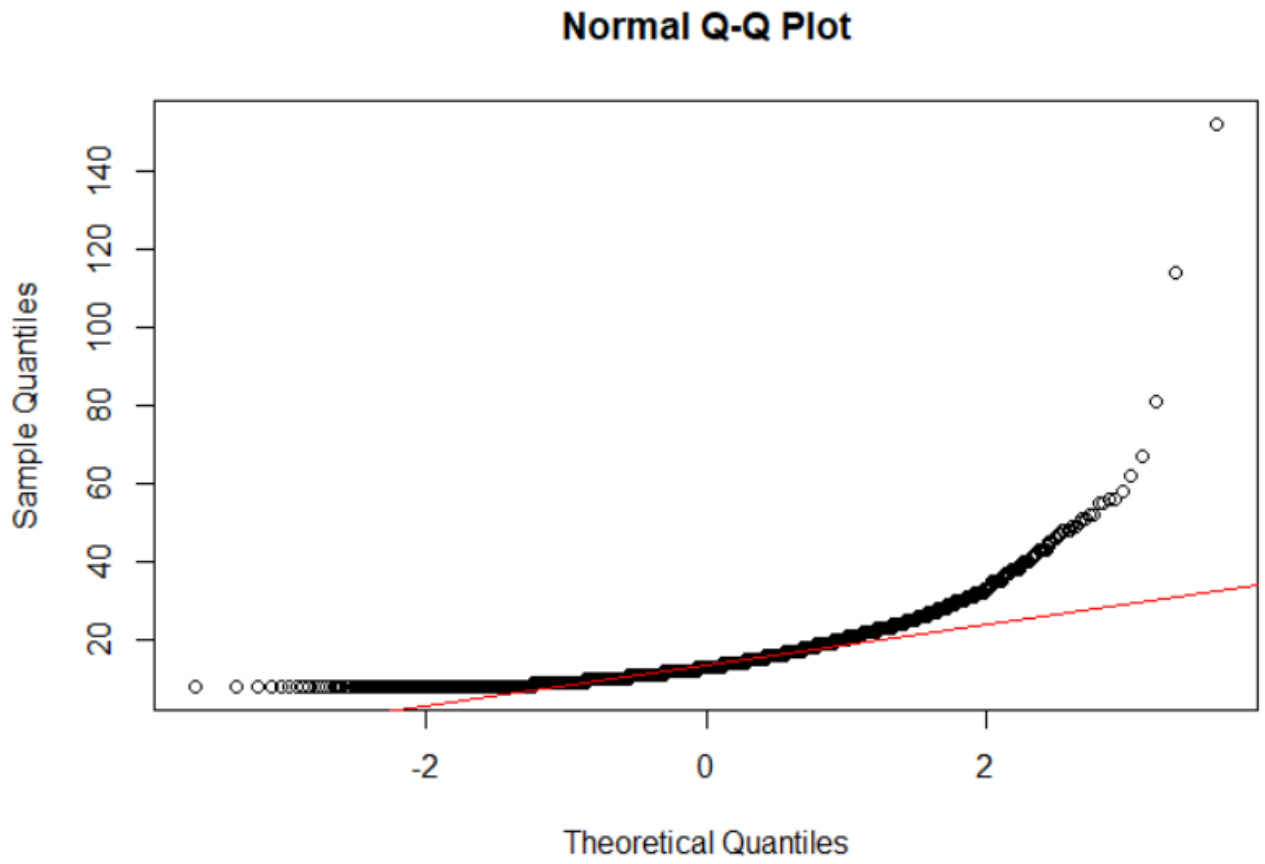
Negative Binomial – all results

Negative Binomial Results

	Dependent variable:								
	negbinempty	negbin1	negbin2	negbin3	purchase negbin4	negbin5			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
distance		0.125***	0.124***	0.125***	0.125***	0.125***	0.125***	0.125***	0.124***
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
income		0.015***	0.015***	0.015***	0.015***	0.015***	0.015***	0.015***	0.015***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
cust_age		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
facebookvisit1		0.078***	0.072***	0.077***	0.074***	0.078***	0.077***	0.073***	0.074***
		(0.015)	(0.018)	(0.018)	(0.015)	(0.015)	(0.018)	(0.018)	(0.015)
gender		0.017	0.018	0.017	0.018	0.017	0.017	0.018	0.018
		(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
numoffriends			0.0001	0.00004			0.00005	0.00005	
			(0.0003)	(0.0003)			(0.0003)	(0.0003)	
numofposts			0.0003		0.0003			0.0003	0.0003
			(0.0002)		(0.0002)			(0.0002)	(0.0002)
PublicProfile			0.008			0.008	0.008		0.008
			(0.014)			(0.014)	(0.014)		(0.014)
Constant	2.696***	2.222***	2.160***	2.218***	2.168***	2.219***	2.215***	2.163***	2.166***
	(0.007)	(0.043)	(0.067)	(0.055)	(0.057)	(0.044)	(0.055)	(0.066)	(0.057)
Observations	3,677	3,677	3,677	3,677	3,677	3,677	3,677	3,677	3,677
Log Likelihood	-11,742.730	-11,331.950	-11,330.690	-11,331.940	-11,330.860	-11,331.790	-11,331.780	-11,330.850	-11,330.700
theta	8.766*** (0.308)	12.274*** (0.487)	12.291*** (0.488)	12.274*** (0.487)	12.288*** (0.487)	12.277*** (0.487)	12.277*** (0.487)	12.288*** (0.487)	12.291*** (0.488)
Akaike Inf. Crit.	23,487.460	22,675.900	22,679.380	22,677.890	22,675.730	22,677.570	22,679.550	22,677.700	22,677.410

Note: *p<0.05; **p<0.01; ***p<0.001

Normalization



Poisson- all results

```
> lrtest(poisson1, poissonempty) # possion is not a good fit
Likelihood ratio test

Model 1: purchase ~ distance + income + cust_age + facebookvisit + gender
Model 2: purchase ~ 1
  #Df LogLik Df  Chisq Pr(>Chisq)
1   6 -12416
2   1 -13428 -5 2024.4  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lrtest(poisson2, poissonempty) # possion is not a good fit
Likelihood ratio test

Model 1: purchase ~ distance + income + cust_age + facebookvisit + gender +
    numoffriends + numofposts + PublicProfile
Model 2: purchase ~ 1
  #Df LogLik Df  Chisq Pr(>Chisq)
1   9 -12412
2   1 -13428 -8 2032.2  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lrtest(poisson1, poisson2)
Likelihood ratio test

Model 1: purchase ~ distance + income + cust_age + facebookvisit + gender
Model 2: purchase ~ distance + income + cust_age + facebookvisit + gender +
    numoffriends + numofposts + PublicProfile
  #Df LogLik Df  Chisq Pr(>Chisq)
1   6 -12416
2   9 -12412  3 7.7724    0.05096 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Poisson Results

	Dependent variable:		
	purchase		
	poissonempty (1)	poisson1 (2)	poisson2 (3)
distance		0.123***	0.123***
		(0.003)	(0.003)
income		0.015***	0.015***
		(0.001)	(0.001)
cust_age		0.001	0.001
		(0.001)	(0.001)
facebookvisit1		0.080***	0.075***
		(0.010)	(0.012)
gender		0.018	0.018
		(0.012)	(0.012)
numoffriends			-0.00000
			(0.0002)
numofposts			0.0003*
			(0.0001)
PublicProfile			0.011
			(0.009)
Constant	2.696***	2.227***	2.163***
	(0.004)	(0.029)	(0.045)
Observations	3,677	3,677	3,677
Log Likelihood	-13,427.670	-12,415.490	-12,411.600
Akaike Inf. Crit.	26,857.350	24,842.970	24,841.200
Note:			*p<0.05; **p<0.01; ***p<0.001

Multicollinearity – all variables

	distance	income	gender	facebookvisit	cust_age	numoffriends	numofposts	PublicProfile
distance	1.000	-0.068	0.148	0.364	0.098	0.201	0.080	0.038
income	-0.068	1.000	-0.348	-0.188	-0.098	-0.128	-0.036	-0.014
gender	0.148	-0.348	1.000	0.235	0.026	0.144	0.029	0.012
facebookvisit	0.364	-0.188	0.235	1.000	0.221	0.583	0.206	0.036
cust_age	0.098	-0.098	0.026	0.221	1.000	0.096	0.050	-0.012
numoffriends	0.201	-0.128	0.144	0.583	0.096	1.000	0.107	0.006
numofposts	0.080	-0.036	0.029	0.206	0.050	0.107	1.000	0.012
PublicProfile	0.038	-0.014	0.012	0.036	-0.012	0.006	0.012	1.000

No variable from the 8 input variables has collinearity problem.

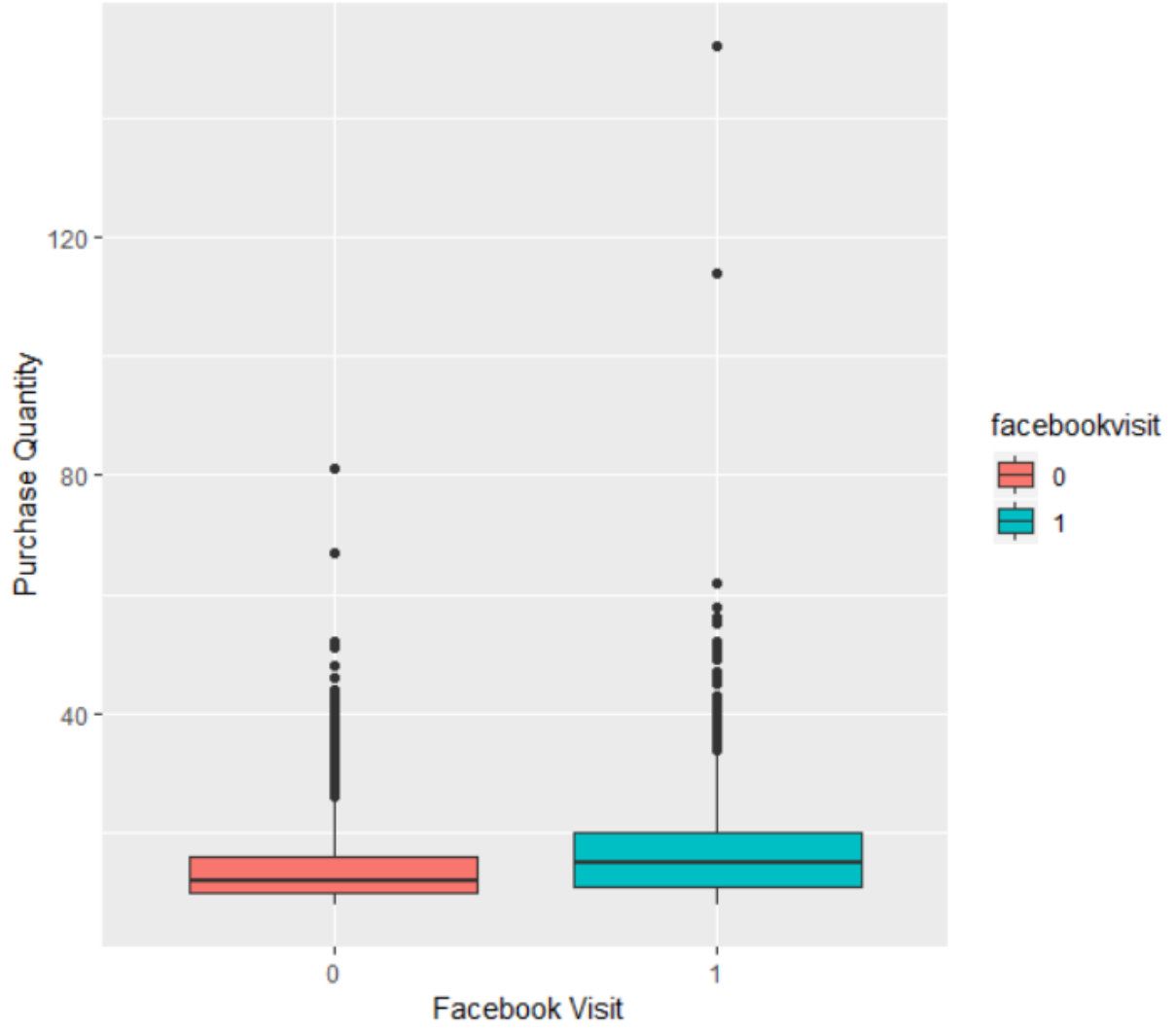
The linear correlation coefficients ranges between:
min correlation (PublicProfile ~ numoffriends): 0.006197103
max correlation (numoffriends ~ facebookvisit): 0.5833385

---------- VIFs of the remained variables --------

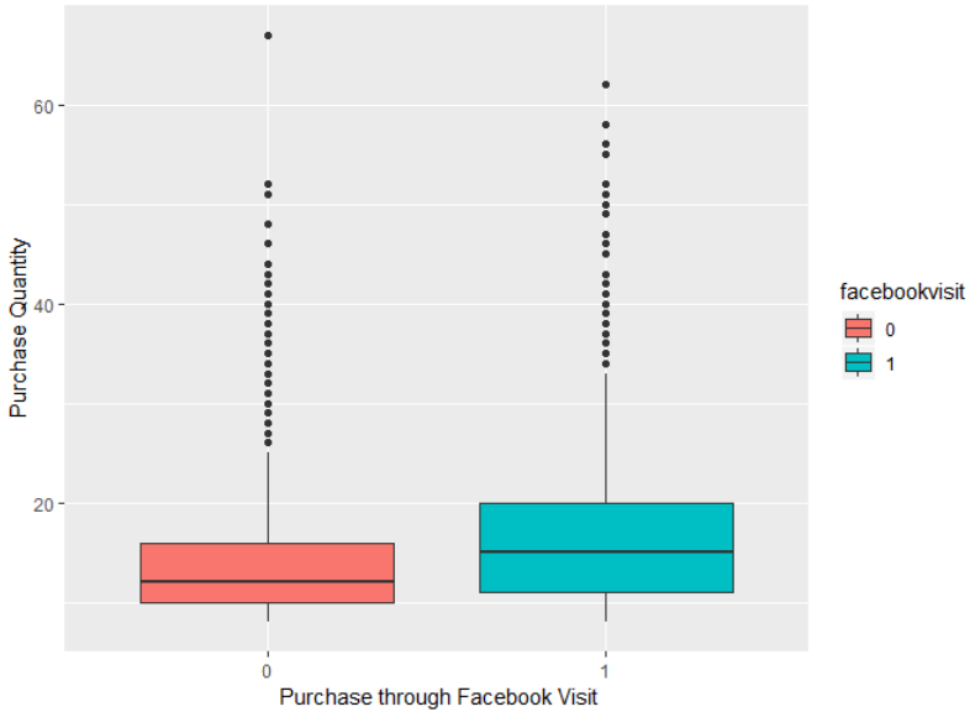
	Variables	VIF
1	distance	1.160624
2	income	1.160494
3	gender	1.187143
4	facebookvisit	1.847091
5	cust_age	1.060565
6	numoffriends	1.520667
7	numofposts	1.045115
8	PublicProfile	1.002961

Outlier Detection

- Outliers present in purchase quantity: 152 and 114
- Both purchases were made after facebook visit



Outlier Detection- Descriptive Statistics




```
 Descriptive statistics by group
group: 0
             vars    n   mean    sd median trimmed   mad min max range
income          1 2450  11.69  3.69     12   12.05  2.97   0  17    17
gender          2 2450   0.10  0.31      0    0.01  0.00   0   1     1
facebookvisit   3 2450   0.00  0.00      0    0.00  0.00   0   0     0
distance        4 2450   1.49  1.18      1    1.41  1.48   0   6     6
cust_age        5 2450  33.25  5.99     32   32.87  7.41  25  50    25
purchase        6 2450  13.78  5.84     12   12.86  4.45   8  52    44
numoffriends    7 2450  99.08 19.85     99   99.10 19.27  31 165   134
numofposts      8 2450 199.67 34.40    199  199.45 34.10  91 309   218
PublicProfile   9 2450   0.27  0.45      0    0.22  0.00   0   1     1
------------------------------------------------------------------------
group: 1
             vars    n   mean    sd median trimmed   mad min max range
income          1 1222  10.16  3.90     11   10.37  2.97   0  17    17
gender          2 1222   0.29  0.45      0    0.24  0.00   0   1     1
facebookvisit   3 1222   1.00  0.00      1    1.00  0.00   1   1     0
distance        4 1222   2.54  1.40      2    2.48  1.48   0   8     8
cust_age        5 1222  36.25  6.65     36   36.45  8.90  25  50    25
purchase        6 1222  16.58  7.74     15   15.43  5.93   8  58    50
numoffriends    7 1222 129.21 19.63    129  129.00 19.27  74 206   132
numofposts      8 1222 215.13 35.89    214  215.32 35.58 100 344   244
PublicProfile   9 1222   0.31  0.46      0    0.26  0.00   0   1     1
```

Outlier Detection- Model

Negative Binomial Results

	Dependent variable:
	purchase
	IRRs
distance	1.1291***
	(0.0048)
income	1.0136***
	(0.0018)
cust_age	1.0016
	(0.0010)
facebookvisit	1.0750***
	(0.0146)
gender	0.9978
	(0.0180)
Constant	9.1937***
	(0.0419)
Observations	3,672
Log Likelihood	-11,189.2100
theta	14.1092*** (0.6066)
Akaike Inf. Crit.	22,390.4200

Note: *p<0.05; **p<0.01; ***p<0.001

Purchase quantity is ***7.5%*** more when the purchase is made though Facebook vs when it is made directly from Amazon.