Life\$tyle: Digitalized New Normal

A Case Study on the conception of consumer behaviour in light of the Paradigm shift in Lifestyle of the Digital Era

> By Soumyajoy Kundu

C.U. Roll Number: 203146 - 21 - 0045

C.U. Registration Number: 146 – 1111 – 0495 – 20

Department of Statistics Maulana Azad College

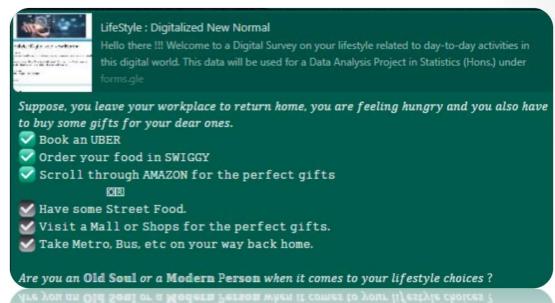
INTRODUCTION

Digitalization is the process of converting existing processes into digital forms.

After the emergence of the internet, digitalization has set its foot in the world.

India is also influenced by the impact of digitalization and hence the government has adopted the "Digital India" campaign in the year 2015.

Nowadays, we can't even imagine a moment without being digitalized. To live a happy and healthy lifestyle, we have some basic criterions to fulfill such as, good friends, quality time with our dear



ones and a bit of travelling and entertainment. What else ?? Now the interesting part is that Digitalization has gifted us with more than what we need within few years and it's upgrading day by day.

This research work focusses upon how much we are digitalized nowadays. It's just a matter of time to wish what we need, and we have it in our fingertips in this Digitalized New Normal.

Are we being controlled by digital world? The purpose of the project not only has a well-built economic prospective but also an imperishable psychological prospective for human kind. Though we love to use our divine gift & being a bit lazy, but still our heart wants to be relive the retro moments. This study portrays the opinions of this Digital Era of being inclined to the fruits of digitalization or to stick with what we are brought up with.

Let's figure out what Statistics and Data Analysis concludes on this interesting topic, "Are you an Old Soul or a Modern Person when it comes to your lifestyle choices?"

OBJECTIVES

This project is an endeavour to figure out the association of one's current status in life and how these affect their experiences while exploring online goods and services platforms.

Here, I have worked with the three most popularly used categories of E-Commerce: Online Shopping, Online Food Delivery and Online Cab-Booking.

For each section, Intra-Categorical analysis has been performed, which consists of statistical tools that fulfill the following objectives:

- 1. To analyze the data set and summarize the composition of each variable according to their respective mentioned categories.
- 2. To test whether there is any significant association between the age, gender, profession, income, frequency of availing the services, of an individual and their opinion regarding preferring digital platforms over the traditional methods.
- 3. To scrutinize how much one's opinion regarding their lifestyle choices is dependent on their status in life.

Along with this, to test the inter-dependency within people's opinions concerning the three different sectors, some techniques for Inter – Categorical Analysis has been used to sum up the project.

PREVIEW OF THE DATA SET

Age Group	Gender	Profession	Where are you situated right now?	Which Income group do you belong to? (Choose the range in which your monthly income belongs) Note: If you are a Student, then select your family monthly income range.
20 to 29	Male	Student	In Kolkata	Rs 75000 to Rs 1.25 lakhs
20 to 29	Female	Student	In Kolkata	Rs 75000 to Rs 1.25 lakhs
20 to 29	Male	Student	Outside Kolkata within West Bengal	Rs 25000 to Rs 75000
20 to 29	Male	Student	Outside Kolkata within West Bengal	less than Rs 25000
20 to 29	Male	Self-Employed	Outside Kolkata within West Bengal	Rs 25000 to Rs 75000

Do you think online shopping is better than traditional offline shopping?	How often have you availed Online Shopping in last 6 months?	the most ?	Which Shopping app do you prefer the most ? (Select Priority Wise) [2nd]	Which Shopping app do you prefer the most ? (Select Priority Wise) [3rd]	What type of products do you usually buy online ? (Select Priority Wise) [1st]	What type of products do you usually buy online ? (Select Priority Wise) [2nd]	What type of products do you usually buy online ? (Select Priority Wise) [3rd]	What is the average price of products you usually purchase?	Rate your Experience in Online Shopping
No	Rarely	Amazon	Flipkart	Others	Electronics	Gift items	Fashion Accessories	Rs 1000 to Rs 5000	4
Yes	Sometimes	Amazon	Flipkart	Myntra	Clothes	Electronics	Gift items	Rs 1000 to Rs 5000	4
Yes	Rarely	Amazon	Flipkart	Myntra	Electronics	Books	Gift items	Below Rs 1000	4
No	Sometimes	Amazon	Flipkart	Myntra	Electronics	Clothes	Footwear	Below Rs 1000	4
Yes	Often	Flipkart	Others	Amazon	Footwear	Clothes	Gift items	Below Rs 1000	4

Which mode of payment do you prefer ? (Select Priority Wise) [1st]	Which mode of payment do you prefer ? (Select Priority Wise) [2nd]	Which mode of payment do you prefer ? (Select Priority Wise) [3rd]	Why do you think people may prefer online facilities (Shopping, Food, Transport) over offline ones? (Select Priority Wise) [1st]	Why do you think people may prefer online facilities (Shopping, Food, Transport) over offline ones? (Select Priority Wise) [2nd]	Why do you think people may prefer online facilities (Shopping, Food, Transport) over offline ones? (Select Priority Wise) [3rd]
Cash on Delivery	Credit & Debit Cards	Net Banking	More Comparing Options	Quality	Cheap
Google Pay	Cash on Delivery	Amazon Pay	Time Saving	Convenient	More Comparing Options
Cash on Delivery	PhonePe	Credit & Debit Cards	Time Saving	Convenient	More Comparing Options
Cash on Delivery	Amazon Pay	Google Pay	Time Saving	More Comparing Options	Cheap
PhonePe	Cash on Delivery	Credit & Debit Cards	More Comparing Options	Time Saving	Convenient

OUTLINE

- 1. Exploratory Data Analysis
 - (i) Pie Chart
 - (ii) Vertical Bar Char (Simple and Grouped)
- 2. Intra Categorical Analysis
 - (i) Chi Square Test for Association
 - (ii) Contingency Tables
 - (iii) Goodman Kruskal's Gamma Measure
 - (iv) Cramer's V
- 3. Multivariate Logistic Regression
- 4. Inter Categorical Analysis

EXPLORATORY DATA ANALYSIS

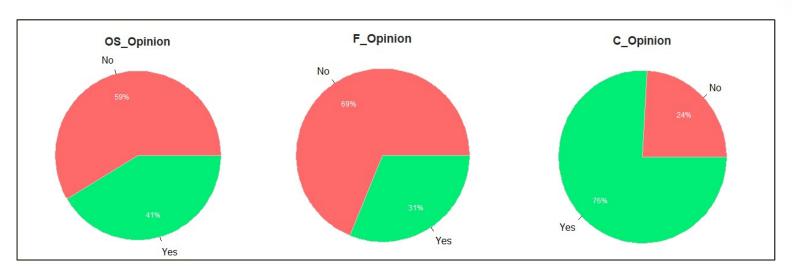


Fig. 1.1: Distribution of Opinions on Online over Offline

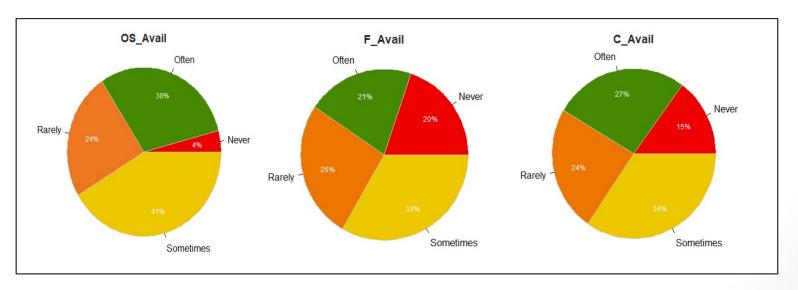


Fig. 1.2: Distribution of Frequency of Availing Online Services in last 6 months

Fig. 1.3: Most Preferred Online Apps of each sector

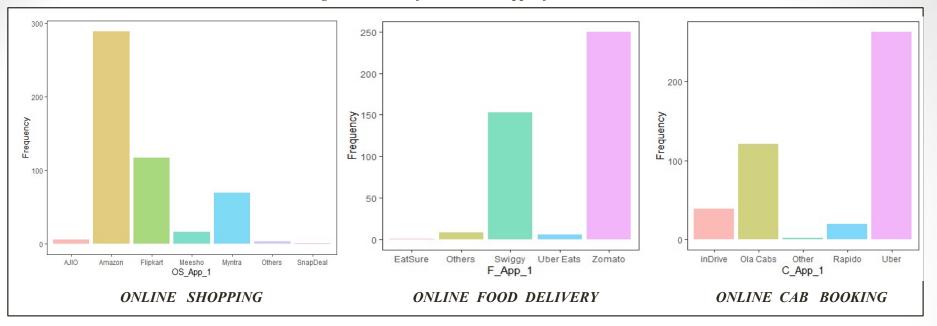


Fig. 1.4: Most Preferred Products and Cuisines

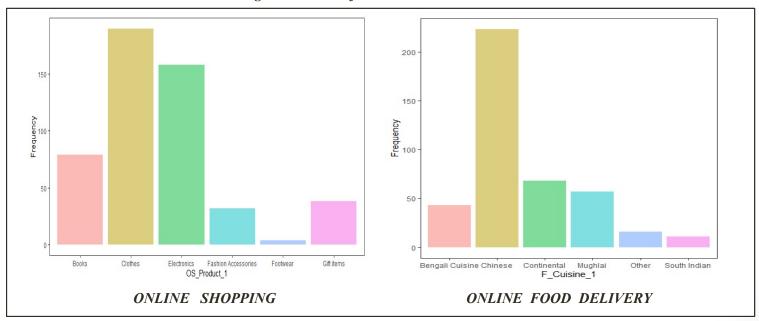


Fig. 1.5: Age vs Gender vs Opinions on each Sector

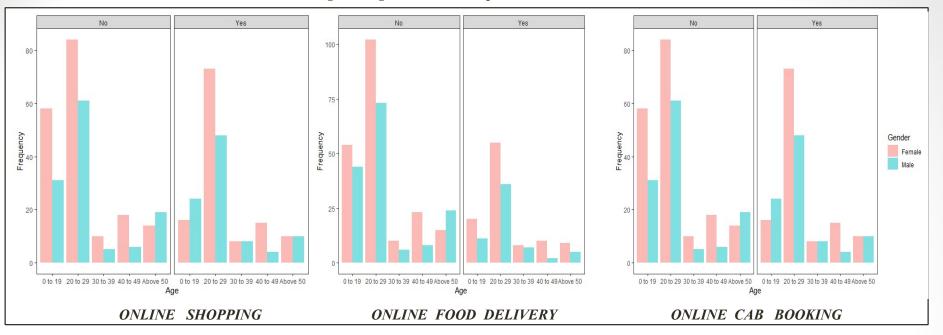


Fig. 1.6: Income vs Frequency of Availing every sector



Here, a preview of the EDA is presented. Likewise, we have analyzed other interesting facts and also concluded for the three most preferred apps, products, cuisines in the report.

ONLINE SHOPPING

- ➤ 41% Yes
- ➤ 4% *Never* Availed
- Amazon, Flipkart and Myntra are the most preferred apps.
- Flectronics, Clothes and Gift Items turns out to be the most demanding products.

ONLINE FOOD DELIVERY

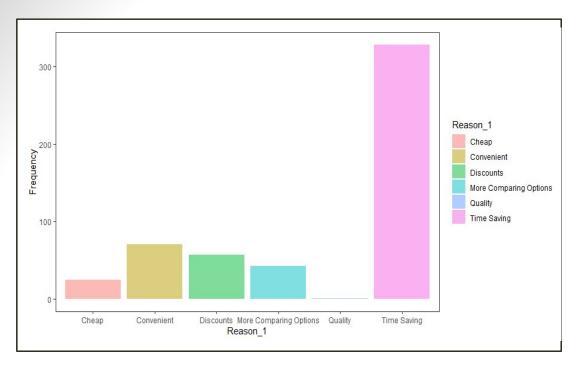
- > 31% Yes
- ➤ 20% Never Availed
- ➤ Zomato and Swiggy are ruling the food-delivery market in India.
- Mostly *all cuisines* are liked when opting for online-food delivery, with *Chinese and Mughlai* winning the foodbattle by a margin.

ONLINE CAB BOOKING

- > 76% Yes
- ➤ 15% *Never* Availed
- Almost all the cab booking apps are explored by people, with an inclination towards Ola Cabs and Uber

Interestingly, contrary to common perception, majority of people would rather go around physically from one shop to another than scroll through the online shopping apps, and would rather have home-cooked meals or go for a dine-out than order in food from their favourite food-delivery apps. When it comes to travelling, a substantial amount of people identify and thus choose the convenience of booking cabs online over the hassle of availing local public transports.

- □ Comparing the frequency of availing the online services in each of the three categories, we can deduce that people are *more comfortable in using Online Shopping than Online Food Delivery or Online Cab Booking*, and thus tend to *use it more often* than the latter.
- Considering the *young adults*, the *females* notably prefer *offline* methods of shopping over online and the *males* share an equal liking for *both* modes of shopping; while in food, an interesting fact is revealed that even *teenagers and young adults*, who are most exposed to the wave of food delivery, are preferring *home-cooked meals* more. Also, it is clear that people of all ages and both the genders unanimously agree on choosing online cab services.
- In Fig. 1.6, we can see that people who are in the *below Rs 25,000* income group, rarely buy things online but people in the above Rs 75,000 group resort to online shopping quite often. Also people earning more than a lakh per month resort to food delivery quite often while people earning less than 25k generally avoid going for it. The income groups have a mostly uniformly distributed responses among all the frequency categories, just people earning more than a lakh per month tend to avail the cab booking services more frequently.



People perceive that using online platforms for shopping, food delivery, or cab booking save a significant amount of time and money (taking discounts into consideration), and those advantages are why people are primarily switching to these services, thus preferring technology over sentiments.

Fig. 1.7: Reason of favouring online services over traditional ones

Most people still prefer handling cash over online methods of payment like netbanking or UPI or payment through cards. This indicates that irrespective of other sectors in life, majority of the population hasn't been able to move on from the age-old process of monetary transactions, and consider it to be the least risk-involved among all the options.

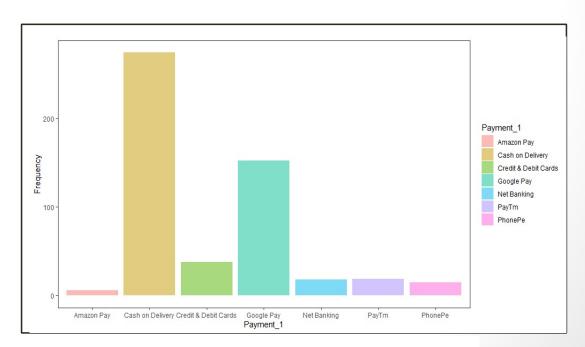
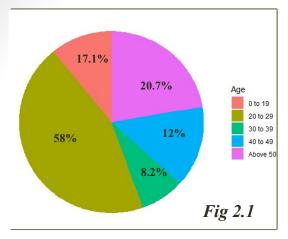
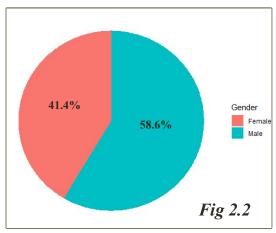
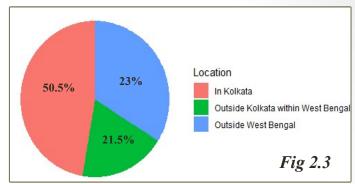


Fig. 1.8: Most Preferred mode of payment

Basic Characteristics of the Surveyed Data

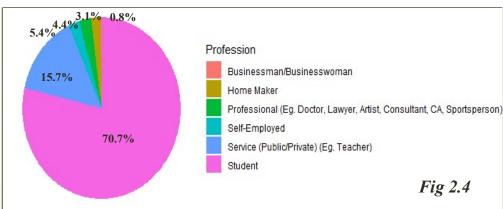


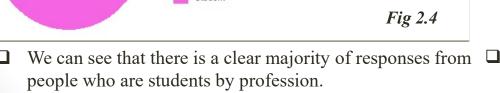


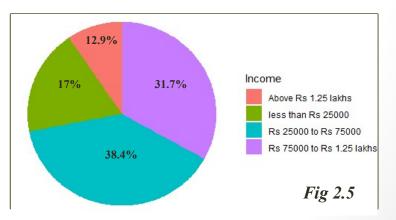


- Although the survey has reached the young adults more, the dataset consists a considerable amount of responses from each age-group, thus helping us to come to an unbiased conclusion.
- Around 58% of the responses are from females, leaving 42% for male responses, thus leading to a near-about equal distribution between the genders, in the data points.

For a survey created solely in Kolkata, we have reached a significant amount of people from outside Kolkata and outside West Bengal as well, which helps us deduce how the opinions of people vary with the place they are currently living in.







Although we have a good count of answers from people belonging to each monthly income group category, people with income ranging from Rs. 25,000 to Rs. 1.25 lakhs have higher representation.

INTRA - CATEGORICAL ANALYSIS

Categorical Data Analysis is the analysis of data where the response variable has been grouped into a set of mutually exclusive ordered (such as age group) or unordered (such as Profession) categories. Categorical (or discrete) variables are used to organize observations into groups that share a common trait.

Contingency Tables & Chi - Square Test for Associations

In statistics, a *Contingency table* (also known as a cross tabulation or crosstab) is a type of table in a matrix format that displays the (multivariate) frequency distribution of the variables.

The *Chi-Square Test for Association* is used to determine if there is any association between two variables. It is really a hypothesis test of independence.

Consider the testing problem,

 H_0 : the two variables are independent vs H_1 : the variables are dependent.

We Reject the null hypothesis at level 0.05 if, p-value = $P(\chi^2 \ge obs(\chi^2)) \le 0.05$

❖ Cramer's V and Goodman – Kruskal gamma

 $\nabla_{AB} = \frac{P_s - P_d}{P_s + P_d}$; P_s : Probability of concordant pair & P_s : Probability of discordant pair

Contingency Tables and Tests for Associations

	OS_Opinion		
Age	No	Yes	Total
20 to 29	145	121	266
30 to 39	15	16	31
40 to 49	24	19	43
Above 50	33	20	53
Below 20	89	40	129
Total	306	216	522

<u>Table 1.1</u> Pearson's Chi-squared test

This Table depicts how Gender is distributed with the opinion of online shopping over offline. Here, p-value > 0.05, thus the null hypothesis is Accepted.

	OS_Opinion		
Income	No	Yes	Total
Above Rs 1.25 lakhs	64	60	124
less than Rs 25000	80	41	121
Rs 25000 to Rs 75000	89	80	169
Rs 75000 to Rs 1.25 lakhs	73	35	108
Grand Total	306	216	522

Table 1.3

Pearson's Chi-squared test

data: T3.1 X-squared = 8.7356, df = 4, p-value = 0.06806

This Table depicts how Age is distributed with the opinion of online shopping over offline. Here, p-value > 0.05, thus the null hypothesis is Accepted.

	OS_Opinion		
Gender	No	Yes	Total
Female	184	122	305
Male	122	94	215
Total	308	216	522

Table 1.2

Pearson's Chi-squared test

data: T3.4 X-squared = 11.364, df = 3, p-value = 0.009912

This Table depicts how Monthly Income Group is distributed with the opinion of online shopping over offline. Here, p-value < 0.05, thus the null hypothesis is Rejected.

	OS_Opinion		
Profession	No	Yes	Total
Businessman/ Businesswoman	1	3	4
Home Maker	15	13	28
Professional	12	11	23
Self-Employed	8	8	16
Service	46	36	82
Student	224	145	369
Grand Total	306	216	522

Table 1.4

Pearson's Chi-squared test

This Table depicts how the rate of availing online shopping is distributed with the opinion of online shopping over offline. Here, p-value < 0.05, thus the null hypothesis is Rejected.

Pearson's	Chi-squared	test	
-----------	-------------	------	--

data: T3.3 X-squared = 2.8763, df = 4, p-value = 0.5787

This Table depicts how Profession is distributed with the opinion of online shopping over offline. Here, p-value > 0.05, thus the null hypothesis is Accepted.

	OS_Opinion		
OS_Avail	No	Yes	Total
Never	19	2	21
Often	37	121	158
Rarely	110	17	127
Sometimes	140	76	216
Total	306	216	522

Table 1.5

Therefore, summarsing all the results above, we note that chi-square test for association has been performed for checking the dependency of one's opinion of choosing online shopping over offline, on variates such as their Age, Gender, Profession, Income, Frequency of availing online services.

Surprisingly, we have found out that the opinions are dependent on Income and Frequency of availing the services and all other co-variates like Age, Gender and Profession do not affect the preference.

❖ Some Interesting facts to verify using Cramers V & GK Gamma

Shopping Apps	Prefer 1	Prefer 2	Prefer 3
Amazon	287	137	51
Flipkart	116	222	86
Myntra	68	81	156
Others	26	57	204

Pearson's Chi-squared test

data: T1 X-squared = 485.86, df = 6, p-value < 2.2e-16

> cramersV(T1) [1] 0.4036488

The table displays the distribution of mostly used apps according to the people's preferences.

Table 2.1

- ☐ From chi-square test of independence, we observe that the null hypothesis is Rejected, i.e. The Shopping Apps and the Order of Proffering them are dependent on each other.
- □ Referring to the table, note that cramer's V turns out to be 0.4036 and degrees of freedom = min(4-1,3-1)=2 indicates a medium (or moderate) assosciation between Order of Preference and mostly used Shopping Apps.

Average Price	Below Rs 1K	Rs 1K to Rs 5K	Above Rs 5K	Total
OS_Avail				
Rarely	73	46	5	124
Sometimes	109	91	12	212
Often	72	78	8	158
Total	254	215	25	494

Table 2.2

Pearson's Chi-squared test

data: T2 X-squared = 5.2344, df = 4, p-value = 0.2641

> cramersV(T2)
[1] 0.07278689

Goodman-Kruskal's gamma

This Table depicts how frequent people use online shopping with respect to their average price of buying products. Note, the above table is an ordinal categorical data.

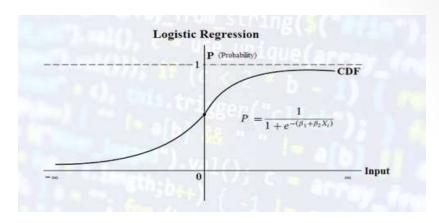
- ☐ From chi-square test of independence, we observe that the null hypothesis is Accepted, i.e. Rate of availing Online Shopping does not depends on the average price of products usually ordered over online.
- Referring to the table, note that cramer's V turns out to be 0.0728 and degrees of freedom = min(3-1,3-1)=2 indicates a small (or weak) association between rate of availing online shopping and the average price of usually bought products.
- ☐ The Goodman-Kruskal's gamma turns out to be 0.1478 which indicates a very weak positive association between the two categories under study.

Likewise, we have analyzed Food and Cab using Chi - Square Test for association and contingency tables. The following conclusions are summarized below:

ONLINE SHOPPING		ONLINE FOOD DELIVERY		ONLINE CAB BOOKI		KING	
Age	A	>	Age	A	>	Age	R
Gender	A		Gender	A	>	Gender	A
Income	R		Income	A	>	Income	R
Profession	A		Profession	A	>	Profession	A
Rate of Availing	R		Rate of Availing	R	>	Rate of Availing	R
 □ The above conclusions are drawn at 5% level of significance. □ In all the three sectors, using Cramer's V, we conclude that there is a moderate association between Order of Preference of people and the most popular apps in each sector, i.e, An individual resort to the most popular apps to get the best service. □ From an ordinal contingency table of Average price of products(or cuisines or trip fare) vs rate of availing the online services, an interesting fact is observed that rate of availing online shopping and booking cab does not depend on the product price or trip fare but on other hand, buying food online gives a low positive association of 0.2887 with the average price of cuisines ordered online. 							

MULTIVARIATE LOGISTIC MODEL

In statistics, the logistic model (or logit model) is a statistical model that models the probability of an event taking place by having the log-odds for the event be a linear combination of one or more independent variables.



* Formulation of the Model based on the surveyed data

We wish to fit a Logistic Model, for what we need a binary response, which corresponds to OS_Opinion (Y) (or F_Opinion or C_Opinion). Here, we have the following co-variates (or explanatory variables),

- \square X_1 : Age
- \square X_2 : Gender
- \square X_3 : Profession
- \square X_4 : Monthly Income Group
- \square X_5 : Rate of Availing Online Shopping in last 6 months

The regression model is given by,

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^{5} \beta_i X_i$$

Three different models corresponding to each sector is analyzed in this study.

Statement of the Problem

In view of our data, we have three major E-Commerce sectors i.e, Shopping, Food and Cab. Here, we wish to scrutinize how much one's opinion regarding their lifestyle choices is dependent on their status in life in each sector. As we have a binary response variable, we have tried to fit Multivariate Logistic Model to analyze and infer about the surveyed data as a whole.

Summary of the Analysis

ONLINE SHOPPING

- ➤ After Significance test, Age and OS_Avail turns out to be significant with atleast one significant sub group.
- $ln\left(\frac{p}{1-n}\right) = \beta_1 X_1 + \beta_5 X_5$

ONLINE FOOD DELIVERY

- ➤ After Significance test, *F_Avail* turns out to be significant with atleast one significant sub group.
- Then the modified model, > Then the modified model, $ln\left(\frac{p}{1-p}\right) = \beta_5 X_5$

ONLINE CAB BOOKING

- > After Significance test, Intercept and *C_Avail* turns out to be significant with atleast one significant sub group.
- Then the modified model, $ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_5 X_5$

Now, we split the data set of 522 data points into Training Data(392 observations) and Testing Data(130 observations), and then fit the above Logistic models using training data. The predictions are made on the testing data and compared with the observed ones.

ONLINE SHOPPING

	Actual		
Predicted	Positive	Negative	
Positive	69	23	
Negative	7	32	

Table 3.1: OS_Confusion Matrix

	AIC	value	= 442	.55
_		1 001 01 0		

$$\triangleright$$
 Accuracy = 0.77

$$\triangleright$$
 Precision = 0.75

$$\triangleright$$
 Recall = 0.91

$$F1 - Score = 0.82$$

$$\rightarrow$$
 AUC value = $\boxed{0.7853}$

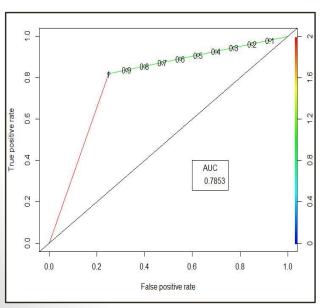


Fig 3.1: OS_ROC Curve

ONLINE FOOD DELIVERY

	Actual		
Predicted	Positive	Negative	
Positive	19	9	
Negative	24	79	

Table 3.2: F Confusion Matrix

- ➤ AIC value = 421.36
- \triangleright Accuracy = 0.75
- \triangleright Precision = 0.68
- ightharpoonup Recall = 0.44
 - F1 Score = 0.53
 - \rightarrow AUC value = $\boxed{0.7228}$

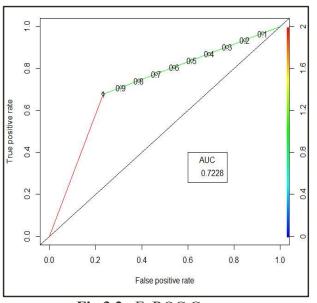


Fig 3.2 : F_ROC Curve

ONLINE CAB BOOKING

	Actual		
Predicted	Positive	Negative	
Positive	69	11	
Negative	20	31	

Table 3.3 : C_Confusion Matrix

- \rightarrow AIC value = 408.38
- \triangleright Accuracy = 0.76
- \triangleright Precision = 0.86
- \triangleright Recall = 0.78
- F1 Score = 0.82
- \rightarrow AUC value = 0.7352

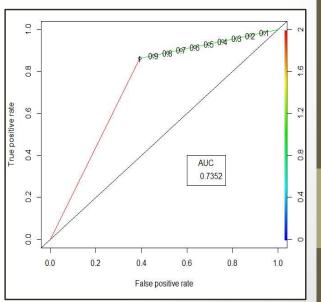


Fig 3.3 : C ROC Curve

INTER - CATEGORICAL ANALYSIS

Here, we have the three different sectors of e-commerce and their opinions of preferring online services over offline. The data is summarized in the table below:

	Yes	No	Total
Online Shopping	216	306	522
Online Food Delivery	163	359	522
Online Cab Booking	366	156	522
Total	745	821	1566

Table 4.1

Statement of the Problem

In view of our data, we have three major E-Commerce sectors i.e, Shopping, Food and Cab. Here, we wish to scrutinize the inter-dependency within people's opinions concerning the three different sectors. Using frequency chi-square technique, we wish to test whether there is any significant difference in the proportion of preferring online services in different sectors.

Consider the tesing problem,

$$H_0$$
: $p_1 = p_2 = p_3 = p_0$

vs

 H_1 : atleast one inequality in H_0

* Test Procedure

Considering all the sectors, we define the test statistic as,

$$T = \sum_{k=1}^{3} \frac{n_k (\widehat{p_k} - \widehat{p_0})^2}{\widehat{p_0} (1 - \widehat{p_0})} \sim \chi^2 \text{ distribution with df 2 (under } H_0)$$

We Reject H_0 at $\alpha (= 5\%)$ level of significance, if the observed value of $T > \chi^2(2,0.05)$

* Analysis

On calculating the observed T, we have obs(T)=170.3066 and $\chi^2(2,0.05)=5.991$, i.e, $obs(T) > \chi^2(2,0.05)$, we Reject H_0 at 5% level of significance.

Also, there are three paired comparisons to test, out of which all are Rejected at 5% level of significance, i.e, all of the e-commerce sectors have significant difference in the proportion of opting Online Services over offline.

! Inference

We can see from the above test procedures that the decisions of people on choosing between E-commerce and their respective offline alternatives, for each of the three categories, are highly co-dependent, which means if one is inclined towards preferring online shopping over offline, there is a high chance that he or she will also prefer the online methods of food delivery and cab booking. This, to an extent, is quite obvious as one who realizes the advantages of one of the three sectors will often tend to use the other two sectors as well.

CONCLUSION

All of us are well aware of the major transformations that E-commerce has brought into our lives, and this project was an effort to estimate the proportion of people preferring the online services over offline ones, and the factors that go behind making their choices. Among Online Shopping, Online Food Delivery and Online Cab Booking, contrary to popular belief, we have seen that people still prefer the offline methods of shopping and ordering food over the online ones, whereas booking cabs online as opposed to the hustles of offline transportations have won majority of people's hearts. Amazon, Zomato and Uber emerge as clear winners among all existing apps in their respective fields of shopping, food delivery and cab booking. The most sought – after products while surfing through online shopping apps are clothes and electronics, while people have a strong affinity towards Chinese cuisine when ordering from food delivery apps. From the Multiple Logistic Regression Analysis, we can infer that age, income and frequency of availing the online services of the person can be labelled as the most significant factors of a person choosing the new digitalised lifestyle over their traditional old habits.

In these hustling times, where each minute and each penny in people's lives are considered to be of supreme value, the data shows that the time saving aspect and the discount of online services are what attracts most customers towards them. On the other hand, if we consider the economic perspective of this, it also proves the growing demand of these platforms. Thus, the market of E – commerce will, in most probability, grow exponentially with time. On the other side of the coin, we have also seen that when it comes to mode of payment while using these services, majority of people prefer cash on delivery over online payment options like UPI or Net banking, which means most people are still not comfortable with monetary transactions on online platforms. Also, only people earning above Rs 50,000 monthly are prone to availing online services often. This shows that there will always be this balance in the population between a group of people who prefer the ease and luxury that comes with technology and another group who believe in holding on to their sweet old habits.

Thank You...

