Amazon Order History

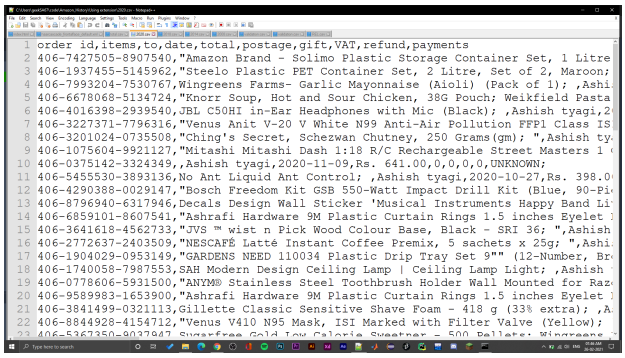
Shantanu Tyagi (aka Shanu)*
A-206 Shiv Shakti Apartments, Plot no. 94, CGHS,
Near AIT Chowk, Sector 54, Gurugram, Haryana, India

In this report, I have scrapped the Amazon order history from 2014 - Present of Mr. Ashish Tyagi's account and analysed the data to understand his family's online spending habits which will keep updating with time. The content shown in this report is the intellectual property of his family and the reader must not reuse, republish or reprint this content without our written consent. The information given is for educational purposes only and I cannot guarantee that it is error free. Thus acting upon this information is at the readers own risk. Do conduct your own diligence before making any financial decisions based on this report since this report is not a substitute for a professional advise. I am not affiliated to Amazon.com in any way and I am not liable for any loss or damage that may arise as a result of your use of this information.

I. INTRODUCTION

The project is divided into three stages as mentioned below:

Scrapping - Since Amazon.com discontinued regular users downloading their order history as a *.csv* file, We need to use web scrapping to extract data from the website. Web scrapping is a process of automatically extracting data from a website which allows our code to mimic a web browser and extract useful raw HTML code from the website and storing in our local device in a convenient format unlike displaying that HTML code like web browsers do. After logging in to Amazon.com, the user can view their orders along with date and amount paid. This information is extracted using a web crawler that generated a *.csv* file containing the order history data that can now be analysed in MS Excel or Python.



```
order id,items,to,date,total,postage,gift,VAT,refund,payments
406-7427505-8907540,"Amazon Brand - Solimo Plastic Storage Container Set, 1 Litre
406-1937455-5145962,"Steelo Plastic PET Container Set, 2 Litre, Set of 2, Maroon;
406-7993204-7530767,Wingreens Farms- Garlic Mayonnaise (Aioli) (Pack of 1); ,Ashi
406-6678068-5134724,"Knorr Soup, Hot and Sour Chicken, 38G Pouch; Weikfield Pasta
406-4016398-2939540,JBL C50HI in-Ear Headphones with Mic (Black); ,Ashish tyagi,2
406-3227371-7796316,"Venus Anit V-20 V White N99 Anti-Air Pollution FFP1 Class IS
406-3201024-0735508,"Ching's Secret, Schezwan Chutney, 250 Grams(gm); ",Ashish ty
406-1075604-9921127,"Mitashi Mitashi Dash 1:18 R/C Rechargeable Street Masters 1
406-0375142-3324349,,Ashish tyagi,2020-11-09,Rs. 641.00,0,0,0,0,UNKNOWN;
406-5455530-3893136,No Ant Liquid Ant Control; ,Ashish tyagi,2020-10-27,Rs. 398.0
406-4290388-0029147,"Bosch Freedom Kit GSB 550-Watt Impact Drill Kit (Blue, 90-Pi
406-8796940-6317946,Decals Design Wall Sticker 'Musical Instruments Happy Band Li
406-6859101-8607541,"Ashrafi Hardware 9M Plastic Curtain Rings 1.5 inches Eyelet
406-3641618-4562733,"JVS ™ wist n Pick Wood Colour Base, Black - SRI 36; ",Ashish
406-2772637-2403509,"NESCAFÉ Latté Instant Coffee Premix, 5 sachets x 25g; ",Ashi
406-1904029-0953149,"GARDENS NEED 110034 Plastic Drip Tray Set 9"" (12-Number, Br
406-1740058-7987553,SAH Modern Design Ceiling Lamp | Ceiling Lamp Light; ,Ashish
406-0778606-5931500,"ANYM® Stainless Steel Toothbrush Holder Wall Mounted for Raz
406-9589983-1653900,"Ashrafi Hardware 9M Plastic Curtain Rings 1.5 inches Eyelet
406-3841499-0321113,Gillette Classic Sensitive Shave Foam - 418 g (33% extra); ,A
406-8844928-4154712,"Venus V410 N95 Mask, ISI Marked with Filter Valve (Yellow);
406-5367350-9037947,Sugarfree Gold Low Calorie Sweetner - 500 Pellets; Wingreens
```

FIG. 1: Snippet of the *.csv* file generated after scrapping

Analysis - The data collected can be split year-wise, month-wise or overall. First one is grouping the data of an entire year into one single unit giving us 8 units. This is done for all the years in our range. Second one is grouping data of a particular month for all the years giving us 12 units, one for each month. Third one is simply showing at what price was each order till date bought for.

Visualisation - Overall data can be easily visualised through line plots. Monthly and yearly data can be easily visualised using bar graphs thus making it easier to compare the data.

II. DATA PROCESSING

A. Formatting

Firstly, we need to convert the data into desired format which from string to integer or decimal value. Since scrapping the price of a product also included the rupee symbol, we need to remove it.

```
def preProcess(df):
    df = df.astype(str)
    def remove_rs(value):
        value = value.replace("Rs. ","")
        value = value.replace(",","")
        return value
    df["total"] = df["total"].apply(remove_rs)
    df["total"] = df["total"].astype(float)
    df["postage"] = df["postage"].apply(remove_rs)
    df["postage"] = df["postage"].astype(float)
    df["gift"] = df["gift"].apply(remove_rs)
    df["gift"] = df["gift"].astype(float)
    df["VAT"] = df["VAT"].apply(remove_rs)
    df["VAT"] = df["VAT"].astype(float)
```

B. Knowing the data

Next, we need to get an idea about how our data is distributed and what are the least, highest, mean, etc values.

```
mean = (df['total'].to_frame()).mean()
tax = df['postage'] + df['VAT'] + df['gift']
print(total.describe())
print(tax.describe())
```

We saw that there was a transaction of of 0 amount done, so that value was removed from the collected data. This gives us the following data description:

*Electronic address: shantanutyagi67@gmail.com

AMOUNT PAID	
count	116.000000
mean	873.795862
std	1625.067723
min	107.100000
25%	270.250000
50%	479.000000
75%	822.750000
max	14999.000000

TAX	
count	116.000000
mean	11.254310
std	22.364908
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	96.000000

III. RESULTS

A. Overall stats

The graph given below shows total price of each of the 116 products that have been purchased till date. The average price has also been plotted which is $\approx Rs.874$. The year in which a particular product was purchased has been shown with a different background color code. We see very few peaks for the recent years which is a result of purchasing essentials on a regular basis. However, during the earlier years the graph has lot of peaks which might denote more expenditure on luxuries.

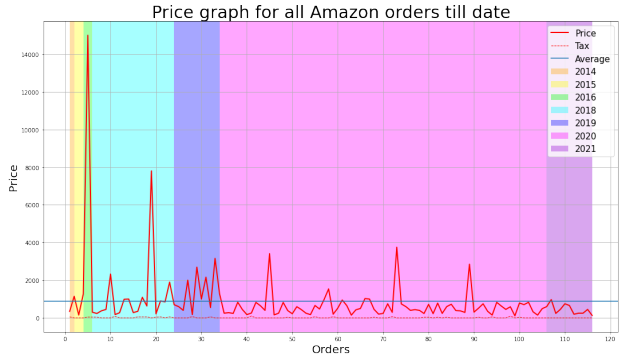


FIG. 2: Price VS Product No.

The graph given below shows tax paid for each of the 116 products that have been purchased till date. The average has also been plotted which is $\approx Rs.11$. The year in which a particular product was purchased has been shown with a different background color code.

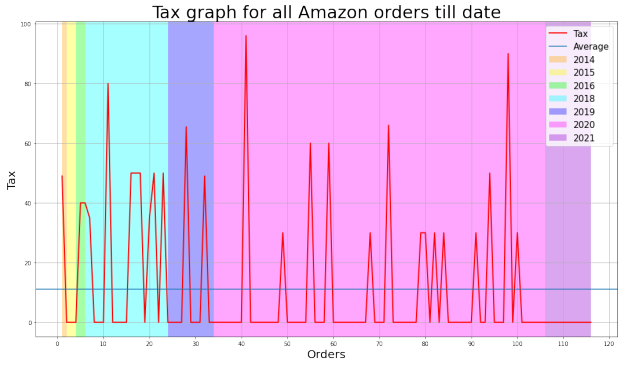


FIG. 3: Tax Paid VS Product No.

B. Yearly stats

The graph given below shows total price, total products purchased, and average price for each year that we have considered. The maximum number of products(72) were purchased in 2020 thus resulting in the highest net expenditure in this year of $\approx Rs.44,300$. However in terms of average expenditure, 2016 tops the list since less quantity of expensive products were bought in that year while the remaining years have an average an average expenditure of around $Rs.7,000$.

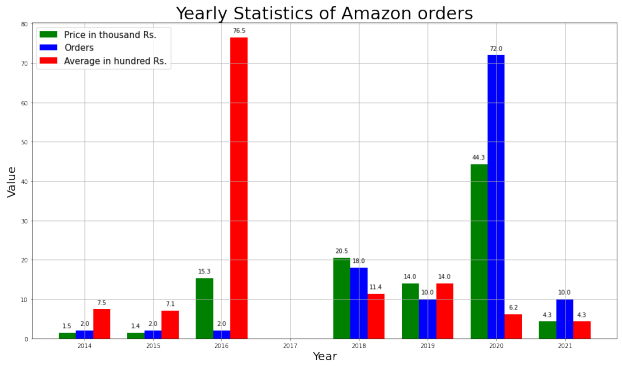


FIG. 4: Yearly Expenditure

C. Monthly stats

The graph given below shows total price, total products purchased, and average price for each month over all the years that we have considered. The maximum number of products(≈ 24) were purchased in the months of *July* and *October* which is more than double or even triple the number of products purchased in other months while the number of products purchased in other months remains constant around 7. This also resulted in the highest net expenditure in in the months of October $\approx Rs.33,700$. However in terms of average expenditure, *April* tops the list with $\approx Rs.2,160$. Colder months seem of have less average expenditure than hotter months.

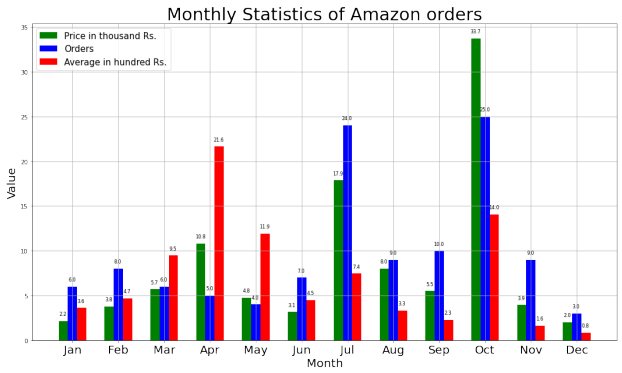


FIG. 5:

IV. CONCLUSIONS

Mr. Ashish Tyagi and his family spends their money wisely with most of the recent purchases being daily necessities or home decor. With the help of the analysis presented above, hopefully the family will be prepared of more expenditure in the months mentioned in the previous section. As far as the pandemic is concerned, it is likely that the same trends will be followed in 2021 as in the preceding year. Nevertheless, wealth consists not in having great possessions, but in having fewer wants and the the constant average expenditure of recent years is its testament.

[1] stackoverflow.com, tex.stackexchange.com