

Direct Research

CSE (498R)

Prepared By

Md.Muttashim Mishel Mawn

ID: 1520800042

Supervised By

Muhammad Shafayat Oshman

Associate Professor

Electrical and Computer Engineering



Department of Electrical & Computer Engineering

North South University

Dhaka 1229

Summer 2022

DECLARATION

I, hereby, declare that the work presented in this study or thesis report is the outcome of the design and development work performed by me under the supervision of **Muhammad Shafayat Oshman**, Associate Professor, Electrical & Computer Engineering, North South University as a course work of CSE/EEE 498R (Direct Research). We also declare that no part of this report has been taken from other works without reference.

Signature of Students

Md.Muttashim Mishel Mawn

APPROVAL

This Thesis report titled 'Data Scraping for Competitive Market Analysis and Behavior' submitted by Md.Muttashim Mishel Mawn ID: 1520800042 to the Department of Electrical and Computer Engineering, North South University, has been accepted as Direct Research Term Final Report.

Signatures

Muhammad Shafayat Oshman

Associate Professor, ECE

Dr. Rajesh Palit

Associate Professor and Chairman, ECE

Abstract

I have proposed a data analysis based study for a competitive market analysis from our country's online based market. Because of now a days a lot of online sale service provides various product at our door but it's so difficult for a newcomer to establish a sustaining business. Competitive analysis is one of the many aspects of digital marketing. It also requires data from various dynamic website. The main objective of our study is to develop a crawler to gather data to compare them and bring up a result that shows us the market status, which type of product they stock most and the demand of a specific product. And for this project I am going to use python library bs4 and scrapy framework for data scraping after that for data filtering I am going to use python pandas. The outcome of this project will help the new comers in the market to understand market status and consumer also can find out the availability of a certain product and what are their services according to specific brands and products.

Table of Content:

Title	Page No
Abstract	3
Table of Content	4
Introduction	5
Problem Statement	5
Description	6
Methodology	6
Technology	11
Discussion	11
Future Work	11
Cost	12
Conclusion	12

Introduction:

In 2022 business is more competitive than 10 years before. Advanced technology gives us the path and the solvent. In Bangladesh business for a newcomer is way harder than the traditional family business man who took the business experience from his father and his father took the business experience from his grandfather. And this is one of the reasons that our new generation is scared to start a new business.

Right now the market is competitive. No one gives you the secret sauce to get success. And I can say competitive market analysis is one of the most important aspects of digital marketing. Maybe everyone can get the initial success by following basic steps of starting a business but what about sustaining business development. If we talk about sustaining business development we should focus on product analysis and need to read consumer philosophy and their capability.

Problem Statement:

Nowadays the competitive market is not newcomer friendly. A lot of newcomers just started their first business but that cannot even sustain. The reason behind this is not enough knowledge about market status, product analysis and consumer psychology to buy a product from you. My approach is to develop a crawler that will scrape data from various ecommerce websites and compare products by price range, demand and supply ratio, services that sellers provide us and how often consumer's write a review. If we are able to compare this we can get the data of the number of stock products in a specific price range. So we can determine the demand of a specific product. After all, it will help newcomers and even the experienced business man too.

Description:

A competitive analysis is a strategy that involves researching major competitors to gain insight into their products, sales, and marketing tactics. Implementing stronger business strategies, warding off competitors, and capturing market share are just a few benefits of conducting a competitive market analysis.

Data scraping involves pulling information out of a website and into a spreadsheet. To a dedicated data scraper, the method is an efficient way to grab a great deal of information for analysis, processing, or presentation.

People can analyze data by themselves but it's not efficient to do manually in 2022. So what if a bot does the whole work and gives you the outcome and does not consume more time? So I want to create a crawler that crawl data from various websites and gathers data in a single csv file. Here you can gather data as much as you can. More data will give you a more appropriate outcome in any situation. The main difference between an experienced businessman and newcomer is the experienced businessman has better knowledge about product and consumer. Even newcomers know those facts but data changes over time so everyone should change their strategies over time too. And this is the main weak point for newcomers. If you fail to adopt you can't survive long. For example: when the rainy season has come the number of umbrella sales increases drastically everyone knows it but it does not depend on the rainy season it depends on how often rain comes. So if you stock a pile of umbrella at the beginning of the rainy season, then it can be said that you have adopted such a strategy due to lack of data analysis of your previous year.

Data and statistics will predict the future. So in this competitive market if you want to survive there is no other way then market and product analysis.

Methodology:

My approach to solving this problem is to break the problem down into smaller parts and solve them one by one. The first thing that I need is information. So I decided to scrape data from the top five computer

hardware and peripheral shops in Bangladesh and select five types of products from their website. Collecting data from various websites is not that much hard but when I crawled data from websites it will take so much unnecessary stuff and it will prevent our next step.

1709	RAPOO E6700 BLUETOOTH ULTRA-SLIM KEYBOARD WITH TOUCHPAD	3850	COMPUTER VILLAGE	Keyboard	In Stock	3850					
1710	RAPOO X9310 WIRELESS ULTRA SLIM ALUMINUM ALLOY KEYBOARD MOUSE COMBO	3475	COMPUTER VILLAGE	Keyboard	In Stock	3475					
1711	RAPOO V500 PRO BACKLIT USB MECHANICAL GAMING KEYBOARD YELLOW AND BLUE	3300	COMPUTER VILLAGE	Keyboard	In Stock	3300					
1712	RAPOO X8210 WIRELESS MOUSE & KEYBOARD COMBO	2175	COMPUTER VILLAGE	Keyboard	In Stock	2175					
1713	TEAM VULCAN Z 8GB DDR4 2666MHZ DESKTOP RAM	4400	Skyland	Ram	In Stock	4400					
1714	TEAM ELITE PLUS 16GB DDR4 2400MHZ DESKTOP RAM	5400	Skyland	Ram	In Stock	5400					
1715	RAMSTA 4GB DDR4 2400MHZ DESKTOP RAM	2300	Skyland	Ram	In Stock	2300					
1716	CORSAIR VENGEANCE LPX 16GB DDR4 3200MHZ DESKTOP RAM	8999	Skyland	Ram	In Stock	8999					
1717	TEAM ELITE U-DIMM 4GB 1600MHZ DDR3 RAM	2100	Skyland	Ram	In Stock	2100					
1718	G.SKILL TRIDENT Z ROYAL 8GB DDR4 3200MHZ SILVER HEATSINK DESKTOP RAM	8500	Skyland	Ram	In Stock	8500					
1719	G.SKILL TRIDENT Z ROYAL SERIES 16GB 3600MHZ RGB SILVER DDR4 RAM	6200	Skyland	Ram	In Stock	6200					
1720	CORSAIR VENGEANCE RGB PRO SL 16GB DDR4 3200MHZ RAM WHITE	8000	Skyland	Ram	In Stock	8000					
1721	G.SKILL TRIDENT Z NEO RGB 16GB DDR4 3600MHZ GAMING DESKTOP RAM	8300	Skyland	Ram	In Stock	8300					
1722	GIGABYTE AORUS RGB 8GB DDR4 3333MHZ DESKTOP GAMING RAM	5200	Skyland	Ram	In Stock	5200					
1723	AITC KINGSMAN DDR3 8GB 1600MHZ HEATSINK DESKTOP RAM	2950	Skyland	Ram	In Stock	2950					
1724	TEAM ELITE U-DIMM 8GB 2400MHZ DDR4 RAM	3299	Skyland	Ram	In Stock	3299					
1725	CORSAIR DOMINATOR PLATINUM RGB 8GB 3200MHZ DDR4 RAM (WHITE)	6300	Skyland	Ram	In Stock	6300					
1726	THERMALTAKE TOUGHTRAM RGB 8GB 4600MHZ DDR4 DESKTOP RAM	10200	Skyland	Ram	In Stock	10200					
1727	AITC 8GB DDR4 UDIMM 2666MHZ DESKTOP RAM	3672	Skyland	Ram	In Stock	3672					
1728	AITC KINGSMAN 4GB DDR4 3000MHZ DESKTOP RAM	2299	Skyland	Ram	In Stock	2299					
1729	TRANSCEND JETRAM 32GB DDR4 3200MHZ U-DIMM DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1730	TRANSCEND JETRAM 16GB DDR4 2666MHZ U-DIMM DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1731	TRANSCEND JETRAM 8GB DDR4 3200MHZ U-DIMM DESKTOP RAM	3500	COMPUTER VILLAGE	Ram	In Stock	3500					
1732	TRANSCEND JETRAM 4GB DDR4 2666MHZ UDIMM DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1733	TRANSCEND 4GB DDR4 2400MHZ UDIMM DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1734	ADATA XPG HUNTER 32GB DDR4 3200MHZ DIMM DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1735	ADATA 16GB DDR4 2666MHZ DESKTOP RAM	8500	COMPUTER VILLAGE	Ram	In Stock	8500					
1736	ADATA XPG HUNTER 16GB DDR4 3200MHZ DIMM DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1737	ADATA XPG SPECTRIX D60G RGB 8GB DDR4 3600MHZ GAMING DESKTOP RAM	5500	COMPUTER VILLAGE	Ram	In Stock	5500					
1738	ADATA XPG SPECTRIX D60G RGB 8GB 3200MHZ DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1739	ADATA GAMMIX D30 8GB DDR4 3000MHZ GAMING DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1740	ADATA XPG GAMMIX D30 8GB DDR4 2666MHZ BLACK HEATSINK GAMING DESKTOP RAM	Call For Price	COMPUTER VILLAGE	Ram	In Stock	null					
1741	PATRIOT SIGNATURE LINE PREMIUM 8GB DDR4 2666MHZ DESKTOP RAM	3000	TECHLAND	Ram	In Stock	3000					
1742	G.SKILL TRIDENT Z NEO 32GB (2 X 16GB) 3600MHZ RGB RAM	15500	TECHLAND	Ram	In Stock	15500					
1743	CORSAIR VENGEANCE LPX 4GB DDR4 2400MHZ DESKTOP RAM	2500	TECHLAND	Ram	In Stock	2500					
1744	CORSAIR VENGEANCE 32GB (2X16GB) DDR5 5600MHZ DESKTOP RAM (BLACK)	38800	TECHLAND	Ram	In Stock	38800					
1745	CORSAIR VENGEANCE 32GB (2X16GB) DDR5 4800MHZ DESKTOP RAM (BLACK)	29000	TECHLAND	Ram	In Stock	29000					
1746	CORSAIR VENGEANCE 16GB DDR5 4800MHZ DESKTOP RAM (BLACK)	14500	TECHLAND	Ram	In Stock	14500					
1747	THERMALTAKE WATERRAM RGB 32GB (4 X 8GB) DDR4 3600MHZ LIQUID COOLING DESKTOP RAM	43500	TECHLAND	Ram	In Stock	43500					
1748	THERMALTAKE WATER RAM RGB LIQUID COOLING 32GB(4 X 8GB) DDR4 3200MHZ DESKTOP RAM	39000	TECHLAND	Ram	In Stock	39000					

Fig.1 Corrupted Data

So I need to organize all of the data in a uniform format. The more we clean the data the more we will get the accurate outcome. So that data cleaning is the most important section for any kind of data analysis project. After that we need to remove all of the corrupted data from the data list. For this project we are storing product name, product price, product category, product link and product review. If any one of the variables is missing we can call it corrupted data. Because when we will start to filter and compare one with another our project outcome accuracy will decrease. But i didn't edit our main dataset; we just created a fresh new dataset in another file. Finally we have a dataset to work with.

	A	B	C	D	E	F	G	H	I
1	Product Title	Price	Shop Name	Category	Availability	G-Price	Links	Review	
2	PNY XLR8 8GB RGB DDR4 3200MHz WHITE DESKTOP RAM	4300	COMPUTER VILLAGE	Ram	In Stock	4300	https://www.computervillage.com.bd/ram-pxy-xlr8-gaming-ram	There have been no reviews for this product yet.	
3									
4	PNY XLR8 EPIC-X RGB 8GB DDR4 3200MHz DESKTOP RAM	4300	COMPUTER VILLAGE	Ram	In Stock	4300	https://www.computervillage.com.bd/pxy-xlr8-epic-x-rgb-desktop	There have been no reviews for this product yet.	
5									
6	G.SKILL TRIDENT-Z NEO 3600MHz 8GB RGB DDR4 3600MHz	11200	COMPUTER VILLAGE	Ram	In Stock	11200	https://www.computervillage.com.bd/g-skill-trident-z-neo-3600mhz-8gb-rgb-dd	There have been no reviews for this product yet.	
7									
8	TEAM T-FORCE DELTA 8GB DDR4 RGB 2666MHz WHITE DESKTOP RAM	4300	COMPUTER VILLAGE	Ram	In Stock	4300	https://www.computervillage.com.bd/team-t-force-delta-8gb-ddr4-rgb-2666mhz	There have been no reviews for this product yet.	
9									
10	TEAM T-FORCE DELTA TUF 8GB RGB DDR4 3200MHz CL16 GAMING DESKTOP RAM	4400	COMPUTER VILLAGE	Ram	In Stock	4400	https://www.computervillage.com.bd/team-t-force-delta-tuf-rgb-8gb-ddr4-3200	There have been no reviews for this product yet.	
11									
12	TEAM T-FORCE VULCAN Z 8GB 3200MHz DDR4 DESKTOP GAMING RAM	3750	COMPUTER VILLAGE	Ram	In Stock	3750	https://www.computervillage.com.bd/team-t-force-vulcan-z-8gb-3200mhz-ddr4-d	There have been no reviews for this product yet.	
13									
14	TEAM ELITE PLUS RED 8GB DDR4 U-DIMM 3200MHz DESKTOP RAM	3650	COMPUTER VILLAGE	Ram	In Stock	3650	https://www.computervillage.com.bd/team-elite-plus-red-8gb-ddr4-u-dimm-3200	There have been no reviews for this product yet.	
15									
16	TEAM ELITE PLUS RED 8GB DDR4 2666MHz RAM	3600	COMPUTER VILLAGE	Ram	In Stock	3600	https://www.computervillage.com.bd/team-elite-plus-red-8gb-ddr4-2666mhz-ram	There have been no reviews for this product yet.	
17									
18	TEAM T-FORCE ZEUS 8GB DDR4 3200MHz DESKTOP GAMING RAM	3600	COMPUTER VILLAGE	Ram	In Stock	3600	https://www.computervillage.com.bd/team-t-force-zeus-8gb-ddr4-3200mhz-deskt	There have been no reviews for this product yet.	
19									
20	TEAM T-FORCE VULCAN TUF 3200MHz 8GB DDR4 GAMING ALLIANCE DESKTOP RAM	3800	COMPUTER VILLAGE	Ram	In Stock	3800	https://www.computervillage.com.bd/team-t-force-vulcan-tuf-3200mhz-8gb-ddr4	There have been no reviews for this product yet.	
21									
22	PNY XLR8 3200MHz 16GB DDR4 DESKTOP GAMING RAM	6300	COMPUTER VILLAGE	Ram	In Stock	6300	https://www.computervillage.com.bd/pxy-xlr8-3200mhz-16gb-ddr4-desktop-gvryqz	There have been no reviews for this product yet.	
23									
24	PNY XLR8 GAMING EPIC-X RGB DESKTOP GAMING RAM	4300	COMPUTER VILLAGE	Ram	In Stock	4300	https://www.computervillage.com.bd/pxy-xlr8-gaming-epic-x-rgb-desktop-g	There have been no reviews for this product yet.	
25									
26	PNY XLR8 3200MHz 8GB DDR4 DESKTOP GAMING RAM	3100	COMPUTER VILLAGE	Ram	In Stock	3100	https://www.computervillage.com.bd/pxy-xlr8-3200mhz-8gb-ddr4-desktop-gaming	There have been no reviews for this product yet.	
27									
28	ADATA 8 GB DDR4 2666 BUS DESKTOP RAM	3100	COMPUTER VILLAGE	Ram	In Stock	3100	https://www.computervillage.com.bd/adata-8-gb-ddr4-2666-bus-desktop-ram	There have been no reviews for this product yet.	
29									
30	GIGABYTE AORUS DDR5 16GB DESKTOP RAM	15500	COMPUTER VILLAGE	Ram	In Stock	15500	https://www.computervillage.com.bd/gigabyte-aorus-ddr5-16gb-desktop-ram	There have been no reviews for this product yet.	
31									
32	TRANSCEND DDR4 32GB 2666MHz DESKTOP RAM	15000	COMPUTER VILLAGE	Ram	In Stock	15000	https://www.computervillage.com.bd/transcend-ddr4-32gb-2666mhz-desktop-ram	There have been no reviews for this product yet.	
33									
34	TRANSCEND DDR4 8GB 2666MHz DESKTOP RAM	3400	COMPUTER VILLAGE	Ram	In Stock	3400	https://www.computervillage.com.bd/transcend-ddr4-8gb-2666mhz-desktop-ram	There have been no reviews for this product yet.	
35									
36	AORUS RGB MEMORY DDR4 16GB (2X8GB) 3333MHz	10000	COMPUTER VILLAGE	Ram	In Stock	10000	https://www.computervillage.com.bd/aorus-rgb-memory-ddr4-16gb-2x8gb-3333mhz	There have been no reviews for this product yet.	
37									
38	AORUS RGB MEMORY DDR4 16GB (2X8GB) 3733MHz	10000	COMPUTER VILLAGE	Ram	In Stock	10000	https://www.computervillage.com.bd/aorus-rgb-memory-ddr4-16gb-2x8gb-3733mhz	There have been no reviews for this product yet.	
39									
40	TEAM XTREEM 8GB 3200 MHz ARGB DDR4 GAMING RAM	6950	COMPUTER VILLAGE	Ram	In Stock	6950	https://www.computervillage.com.bd/team-xtreem-8gb-3200-mhz-argb-ddr4-gamin	There have been no reviews for this product yet.	
41									
42	TEAM DELTA RGB 8GB 3600MHz DDR4 DESKTOP RAM	4700	COMPUTER VILLAGE	Ram	In Stock	4700	https://www.computervillage.com.bd/team-delta-rgb-8gb-3600mhz-ddr4-desktop-	There have been no reviews for this product yet.	
43									

Fig. 2 Dataset

Now we are in data analysis. First we will see how the product stocks within a certain price range. And to determine the range of this product, I extracted the minimum price and maximum price of the product from my data set and divided the price between them into two ways, one lower price range and the other upper price range.

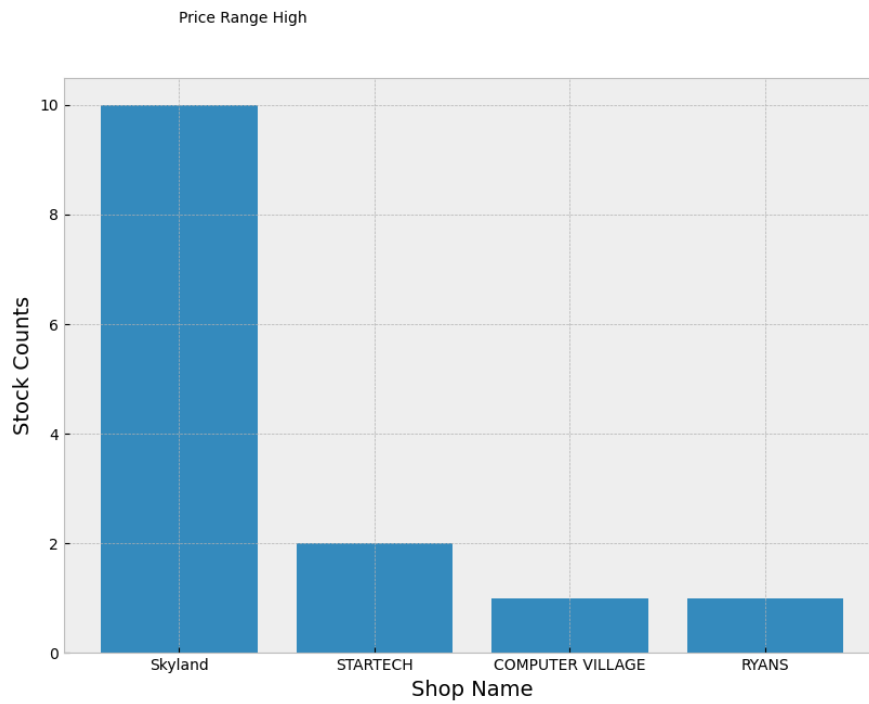


Fig.2 Stock counts in high price range

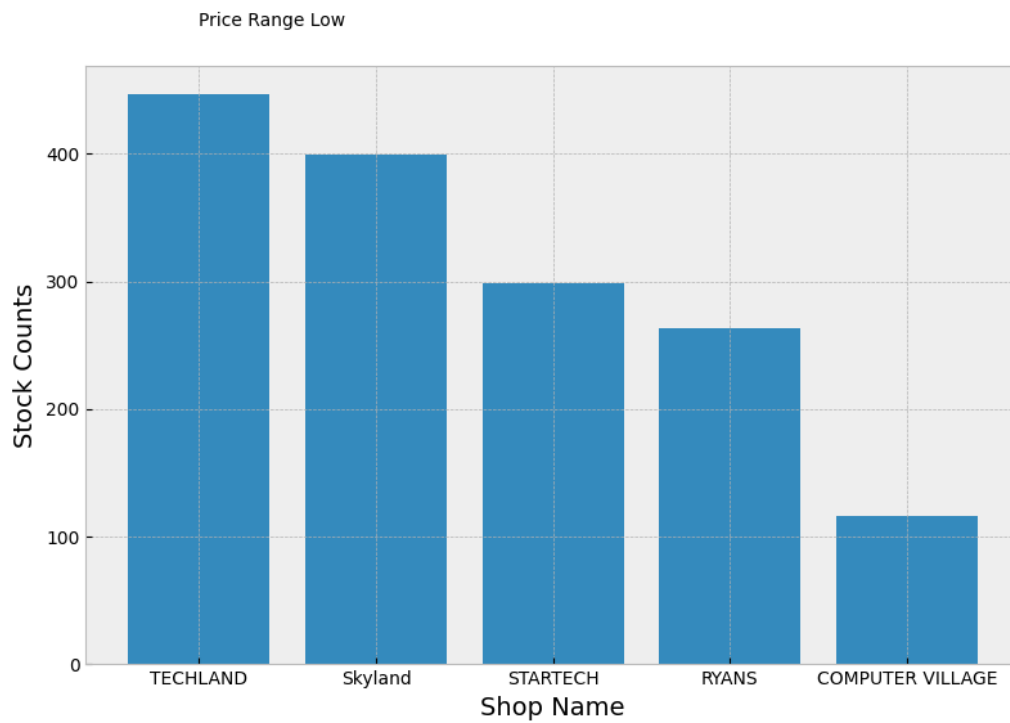


Fig.3 Stock count in low price range

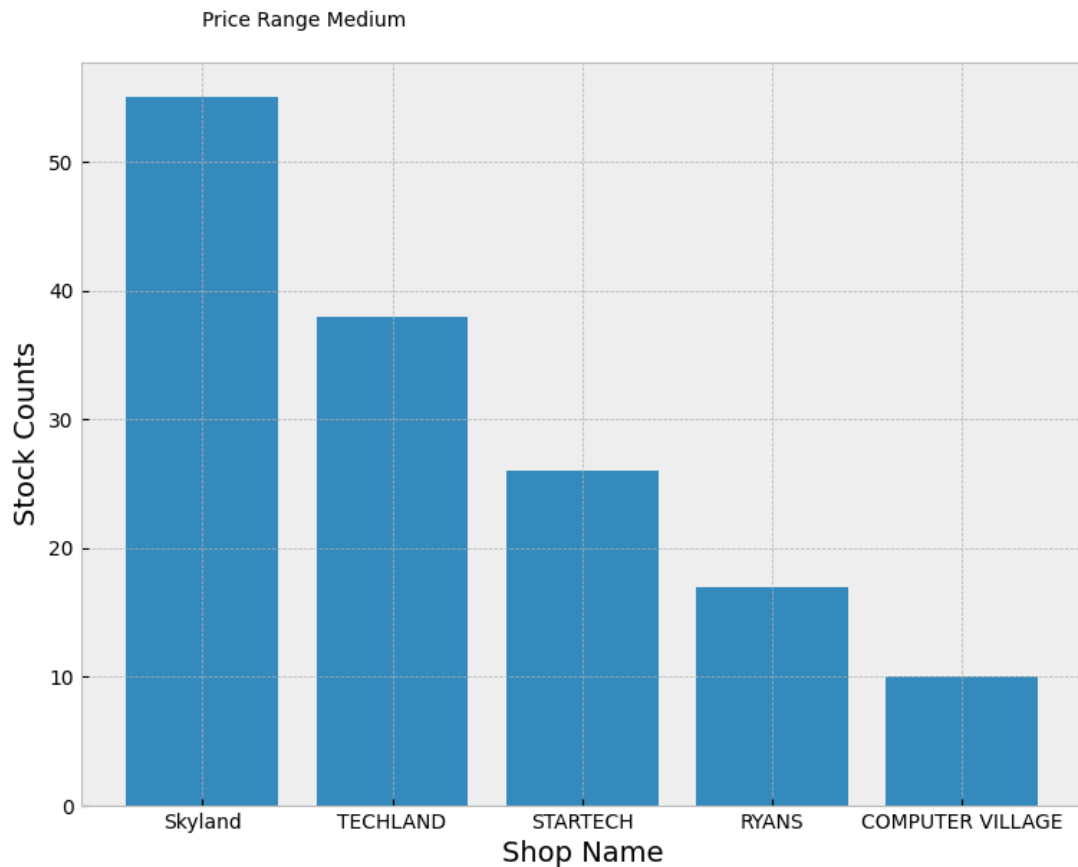


Fig.4 Stock Counts in medium price range

We can understand from the three bar charts that the stocks in the lower price range products are more than the higher price range products. And it is so obvious for any type of market. So we need a specific and more meaningful outcome. And for that i extracted the minimum price range, medium price range and high price range from the data set. And this time I found different results from different product categories. And the outcome will change again over time. Maybe in the future the number of medium price ranges will increase with demand. Because In economics. An increase in supply, all other things unchanged, will cause the equilibrium price to fall, quantity demanded will increase. A decrease in supply will cause the equilibrium price to rise, quantity demanded will decrease.

We have one more variable that will provide us with consumer psychology or trends of any kind of product. If we compare the number of reviews and the number of products within a certain product price range, we will know the consumer's soft spot and which products are more demanded in a specific time period.

Technology:

For this problem I am using python programming language as my base technology. But I am using different libraries for the rest of the work. They are -

- a. Beautiful soup 4
- b. Pandas
- c. Numpy
- d. Matplotlib

Discussion:

Initially I tested 5k data, measured the outcomes and then I tested 10k data and measured the outcomes. And what I saw the result changed when the number of data increased. One more thing I tried during this time period is scraping the same data set in two different time periods but I cannot find any changes because the owner of those websites didn't update their stock and upcoming product information. I can assure that if i can scrape data from two different time periods we will get a more appropriate result.

Future Work:

Basically we have data so we can solve different types of problems too. If we compare individual products in different shops we can find out who provided the minimum price of a specific product. It will help the consumer because comparing a single product from different websites is more time consuming. Also everyone knows that some bad businessmen create a syndicate and they increase a certain product's demand and price by

holding products. So for this problem I need both price data that is specified by the company and the vendor. It will help to break the unnecessary price hike.

Cost:

For this project we just need a better computer to scrape more data in a short time. I scraped 16k data and my computer takes half an hour because I have a mid-end computer. But if we will scrape over 1m data we definitely need high end devices because when i start to scrape multiple pages from a single website i need to put a delay after every page so that server cannot detect it as a bot. If the server detects it, the server will block it. So if we want to avoid this issue we just need a high end device. After that it won't take too much time to execute.

I am working on a computer that costs around 2lakhs. If we want to run this project for industrial purposes we need a computer that is around 5 -7 lakhs. Because we need a faster processing unit.

Conclusion:

In this study I am trying to gather some information about the current market situation that will influence our newcomers to become entrepreneurs. It will not only help the individuals but also puts a huge impact on our central economy, unemployment and a better environment for consumers to buy products.