

Sales Analytics to drive Profitability

A case study of a Fashion E-Commerce Retailer

REVA Academy for Corporate Excellence (RACE)

**Tharuka
Gallekankanamge**

SRN: R19MBA82

Date: 27/08/2022

MBA in Business Analytics

Capstone Project Presentation
Year: II

race.reva.edu.in



- 01. Introduction**
- 02. Literature Review**
- 03. Problem Statement**
- 04. Objectives of Study**
- 05. Project Methodology**
- 06. Business Understanding**
- 07. Data Understanding**
- 08. Data Preparation**
- 09. Modeling**
- 10. Evaluation**
- 11. Analysis and Results**
- 12. Conclusions and Future Scope**
- Bibliography
- Plagiarism Report



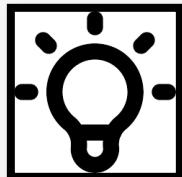
01. Introduction



Ever Changing | Ever Evolving | Ever Challenging



Competitive Landscape



What triggered this study?

Background

- Indian value fashion is an **increasingly fickle** environment
- **Competition** and fashion **heterogeneities**
- **Fast-paced** consumer evolution

Current status

Why this study?

- A **stepping-stone in creating an analytical culture** in the organization.
- Aims to derive **meaningful practical insights** to **increase profitability**
- Introduce statistical **sales forecasting techniques** for better inventory management.



- It's a shark tank!
- Crowded with global and local brands.
- Achieving Profitability with **sharp pricing** is the only way out!

Competitive Landscape



What triggered the study?



- Lack of an analytical approach to data.
- Low budget achievement due to low availability of stock.
- Depleting profitability due to High CODB



02. Literature Review

Let's gather some research around the study!

Business and Industry Related



- **Bigger organizations** have an upper hand due to **more workforce and investment**.
- New-gen companies leverage **emerging tech**.
- India will be the **world's most tech-savvy e-commerce market**
- **Exponential growth** due to the rapid growth of **internet users** in the country



Inventory Management and Sales Forecasting techniques



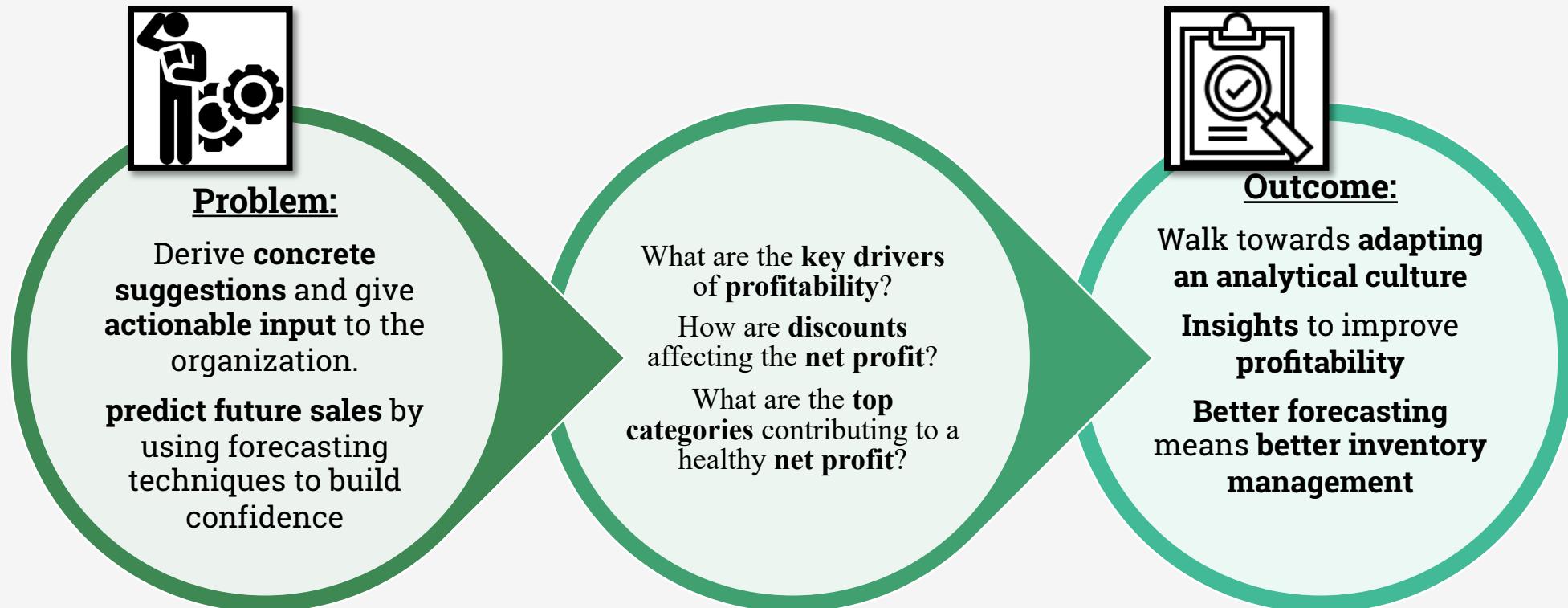
- **Forecasting aspects** that are **unique to the fashion industry** (Seasonality, Large no. of SKUs, Rapid turnaround).
- **Importance of sales forecasting** to secure goods for the future.
- Research on **practical application of ARIMA & SARIMAX** on real business data.



03. Problem Statement

“If you define the problem correctly, you almost have the solution”.

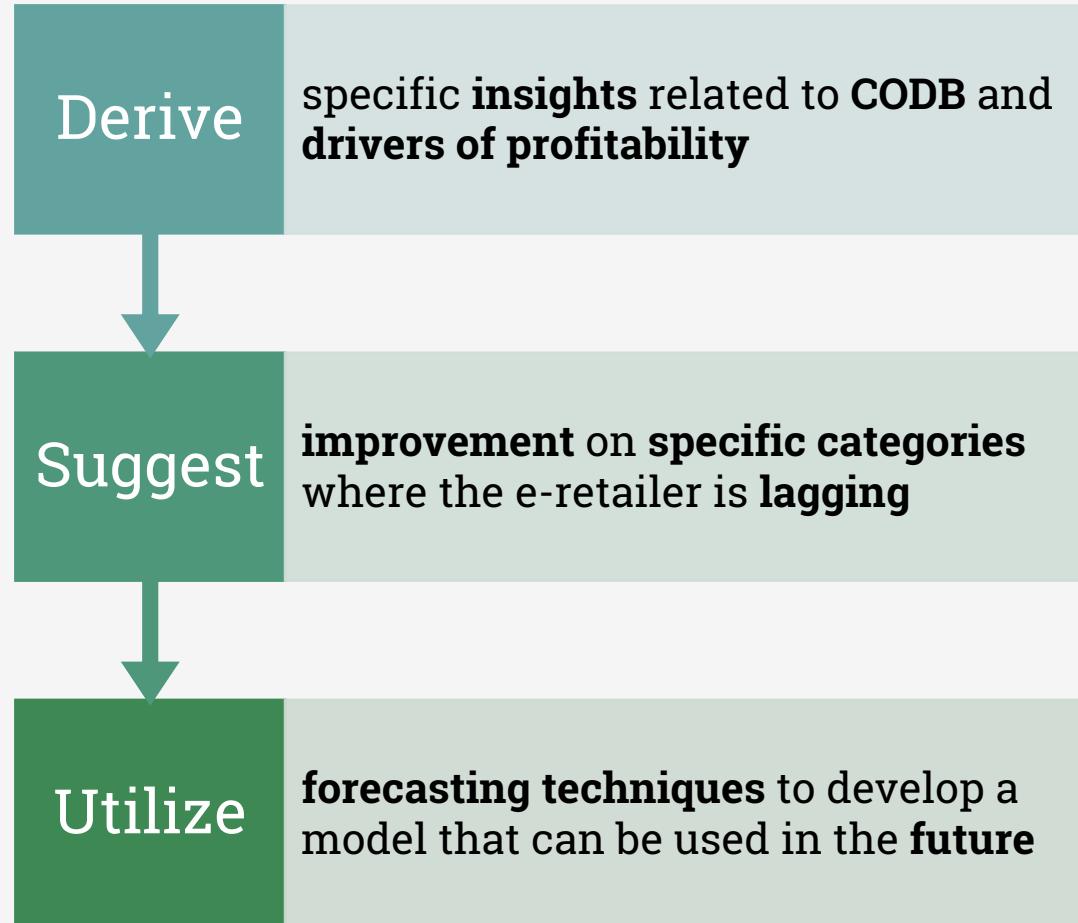
- Steve Jobs



04. Objectives of the study

“No project is completed until its objectives are achieved”.

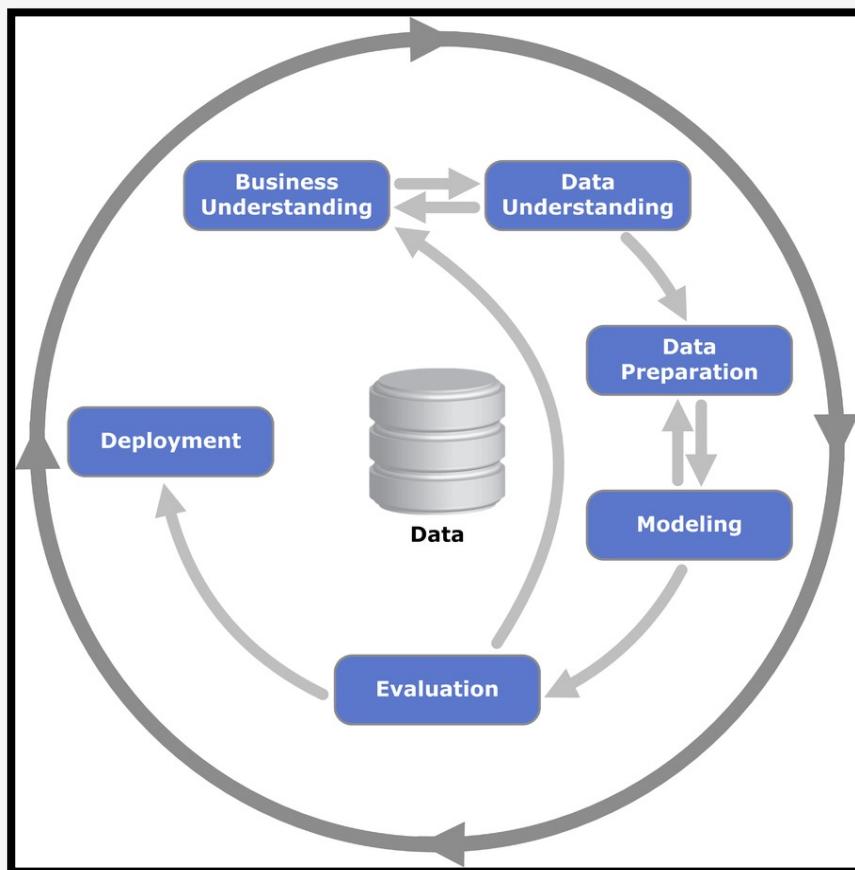
- Paulo Coelho



05. Project Methodology

A framework for data mining and analysis

CRISP - DM



- This study has been conducted as per Cross Industry Standard Process for Data Mining (CRISP-DM) methodology.
- This method has been adopted to derive insights from brand-generated content and used for topic modeling.
- The CRISP-DM methodology is a well-adapted process model across various domains, and it has six phases which is a step-by-step approach to any data science problem.

Figure: Cross-industry standard process for data mining Lifecycle source IBM.

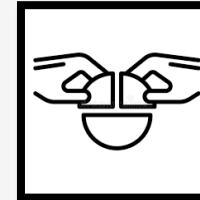
05. Business Understanding

“Understanding the needs of a business is the starting point of any project”.

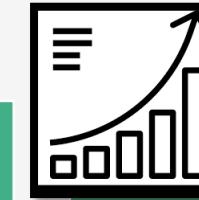
- John Williams



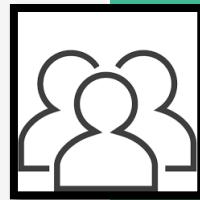
Revenue share of fashion segment: **US\$19.69 bn** in 2022



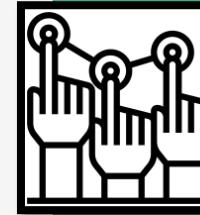
The projected revenue share: **US\$33.11 bn** by 2025



Expected CAGR (2022-2025): **18.92%**



Expected no. of users: **446.2m** users by 2025



User penetration in 2022: 22.8% & hit **30.9%** by 2025



ARPU (The average revenue per user): **US\$61.46**

Source: Fashion E-commerce Report (Statista Digital Market Outlook, 2021)

Dressing The Indian Masses & Unlocking the Fashion Retail

- Online fashion industry continues to grow upwards!
 - Order volume growth of **51%** (FY21)
 - GMV (Gross Merchant Value) growth of **45%** (FY21).



Source: Official websites of above organizations, Fashion E-commerce Report, e-commerce focused supply-chain SaaS technology platform Unicommerce)

“If you don't understand the details of your business you are going to fail”.

— **Jeff Bezos**

Value market segments under Fashion:



Women Apparel



Men Apparel



Kids Apparel



Footwear



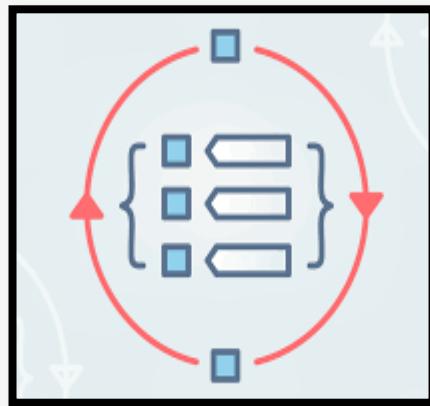
Accessories and bags

Source: Fashion E-commerce Report (Statista Digital Market Outlook, 2021)

07. Data Understanding

“In God we trust, All others must bring data ”.

- W. Edward Demming



Data Source



Data Exploration

You only need a cup of data to make magic.

Raw data
(110 variables)



Cleaned data
(22 variables)



- The dataset is acquired from the organization's OMS.
- Sales of one online channel.
- There were 110 columns in the original data set.
- 2 missing elements – (Geographical data and customer contact details).

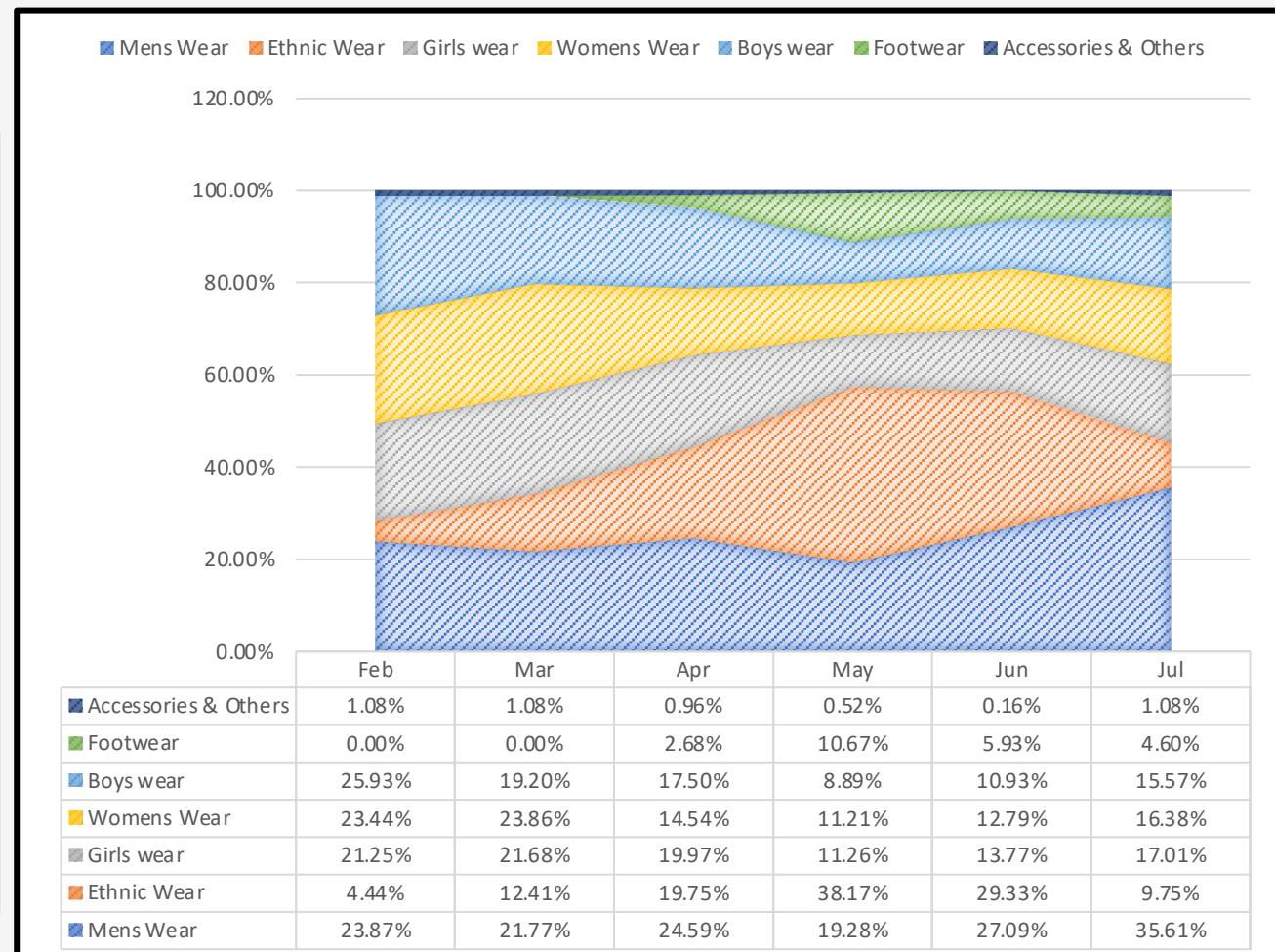
Number of variables	22
Sample size	76008
Missing cells	1749
Missing cells (%)	0.10%
Duplicate rows	80
Duplicate rows (%)	0.10%
Numeric Variables	9
Categorical Variables	13

“Torture the data, and it will confess to anything”.

- Ronald Coase

Category-wise Gross Sales :

- Men's wear:
Stable and the biggest contributor to sales.
- Women's wear:
Dropped since April.
- Kid's wear:
Performed well in the **first 3 months** and **dipped** since **April**.
- Ethnic wear:
A slow start in the beginning and **increase in top line** in May and June.
- Non-Apparel categories:
The contribution is <6%.

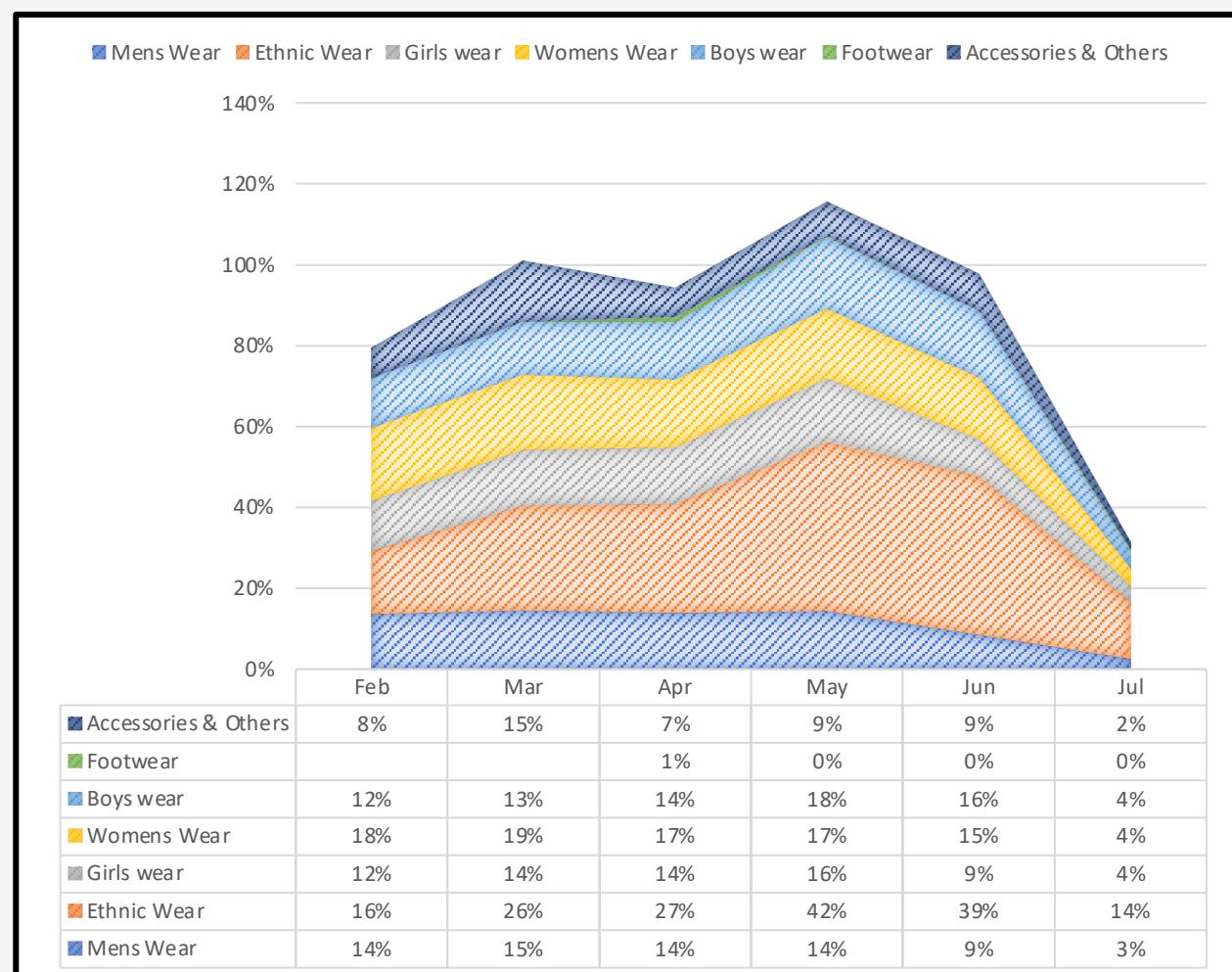


“Torture the data, and it will confess to anything”.

- Ronald Coase

Category-wise Average Discounts Offered:

- Men's wear:
Steady share of business with minimum discounts.
- Ethnic wear:
Topline is achieved via running discounts.
- Other categories:
Maintained a healthy discount %.

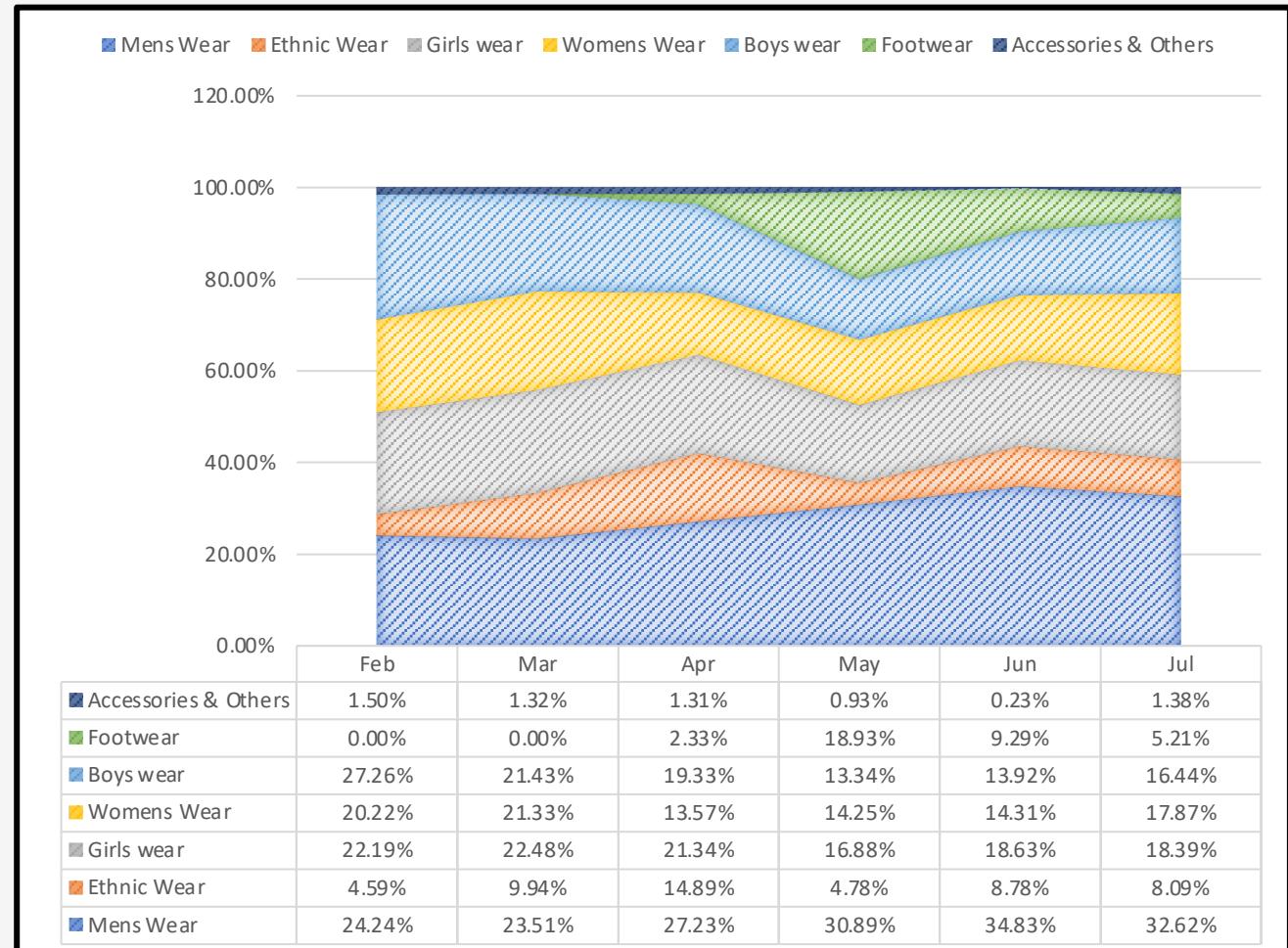


“Torture the data, and it will confess to anything”.

- Ronald Coase

Category wise – Net profit:

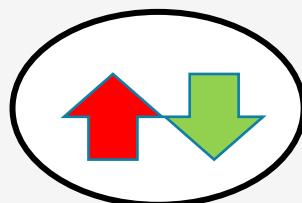
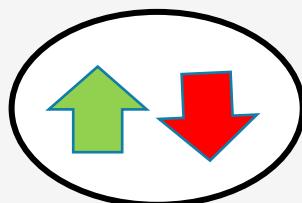
- Ethnic wear:
Least profitable - high discounts run in May and June.
- Other categories:
Healthy Net profit margin.



“Torture the data, and it will confess to anything”.

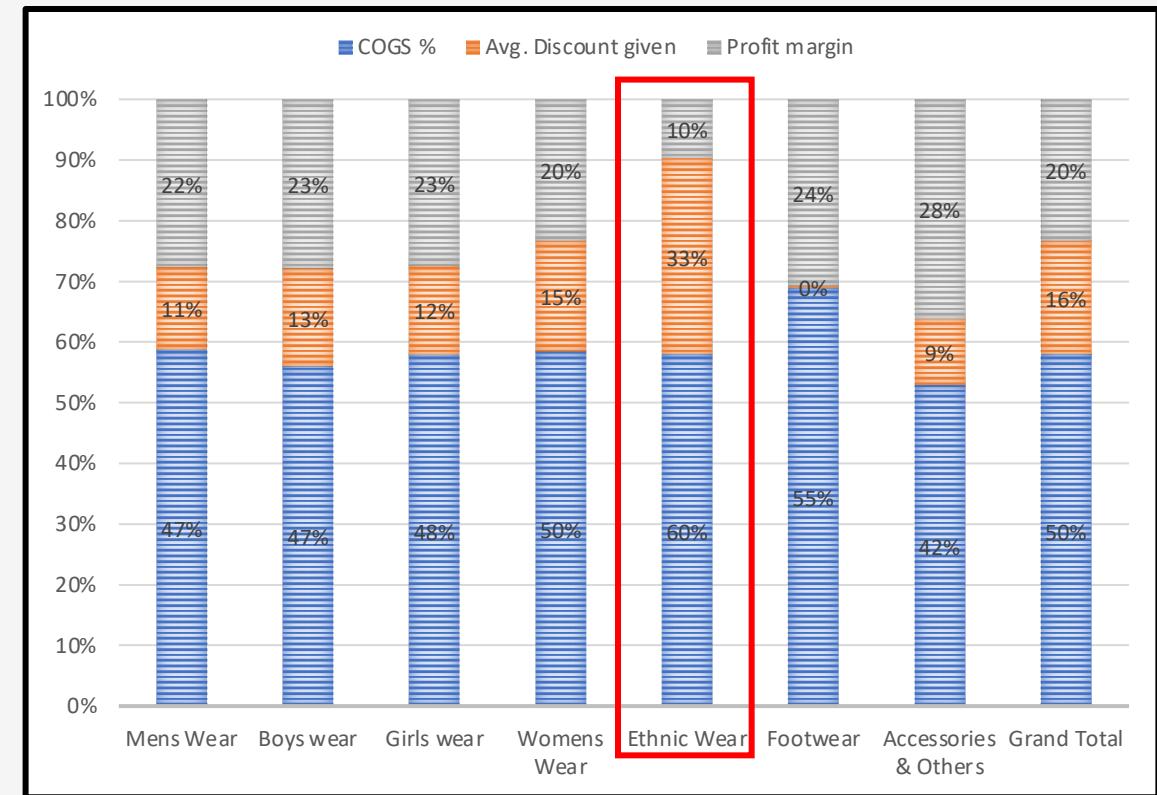
- Ronald Coase

COGS & Disc % & Net Profit %:



$$\text{Profit} = \text{Sales} - \text{Discounts} - \text{COGS}$$

- Ethnic wear is depleting the NP %
- Highest COGS are also ethnic wear.

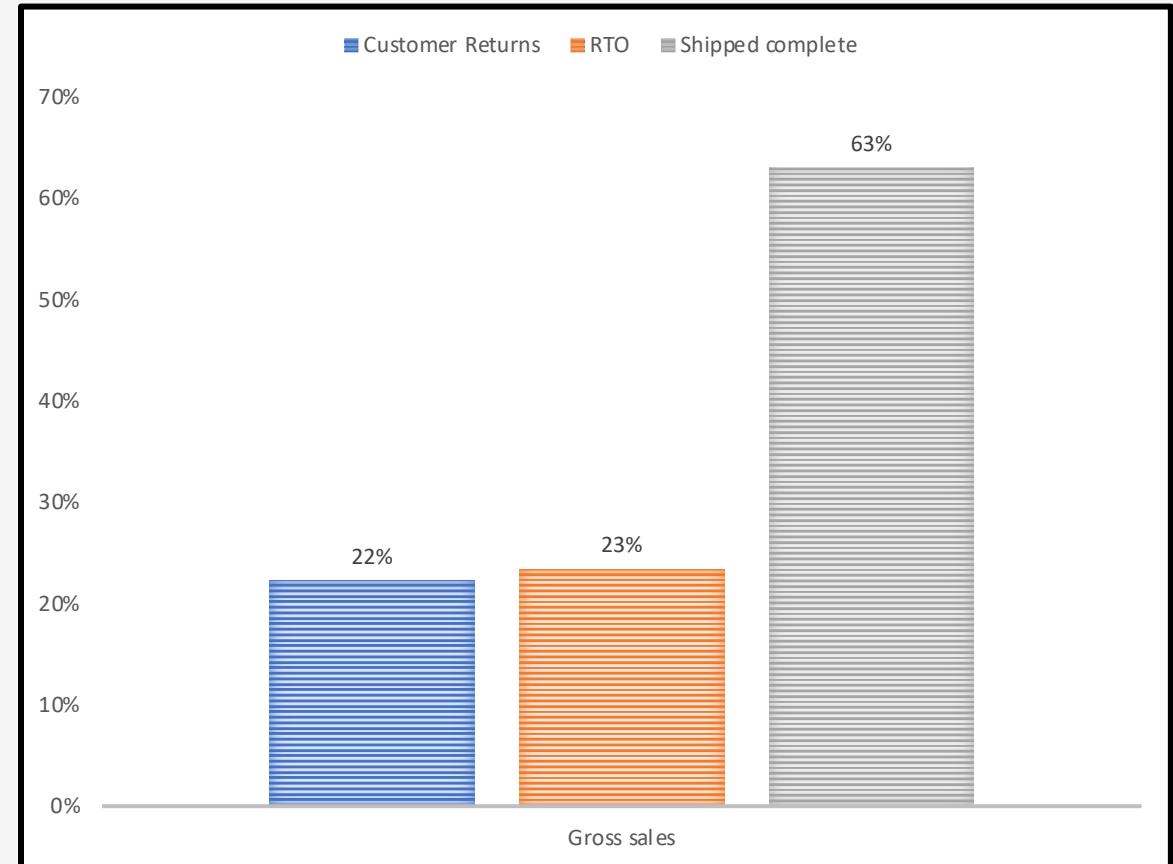


“Torture the data, and it will confess to anything”.

- Ronald Coase

Online Gross sales as per Order Status:

- Final sales impacted the top line is **63% of gross Sales.**
- **45%** of confirmed orders have returned during the past six months.
- Industry standard of <25%.

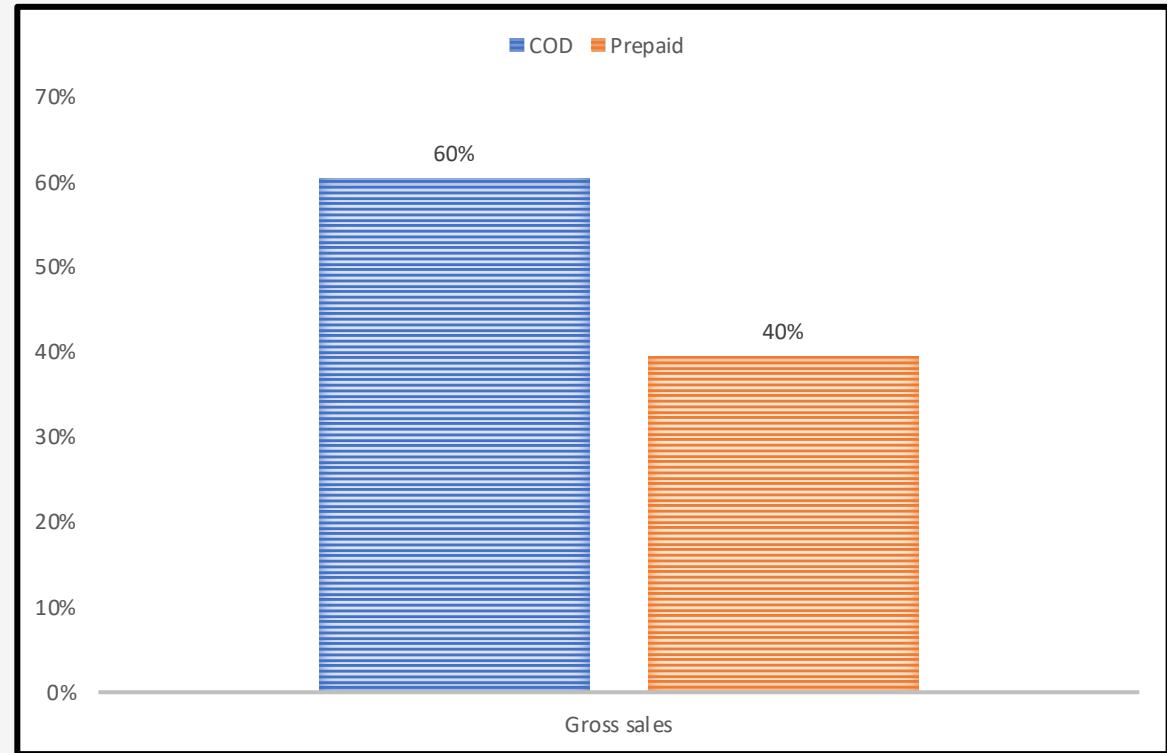


“Torture the data, and it will confess to anything”.

- Ronald Coase

Online Gross sales as per the mode of payment:

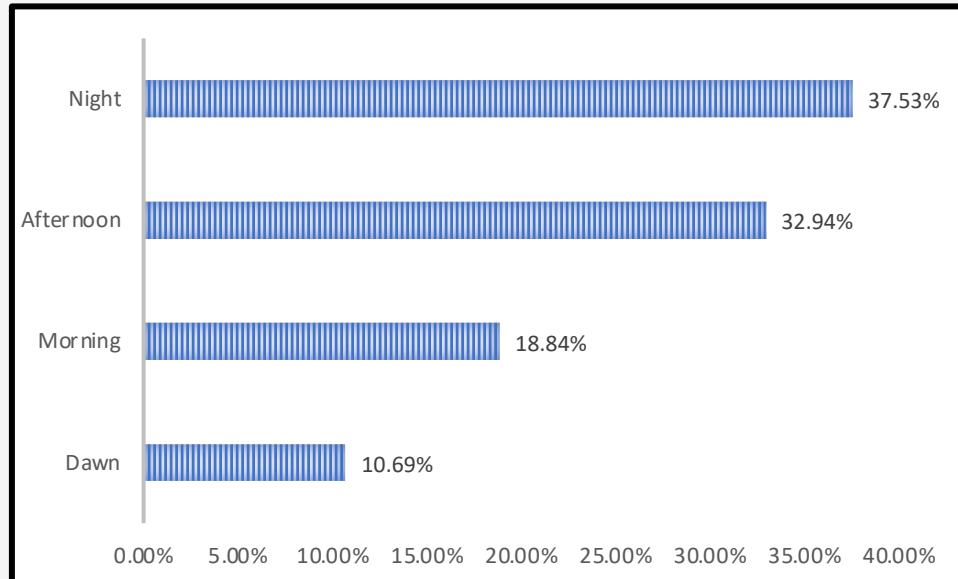
- COD is the **most preferred** mode of payment
- Room for **cancellation / change of mind or circumstances**.
- **Ease of payment** to the delivery partner via UPI payments.



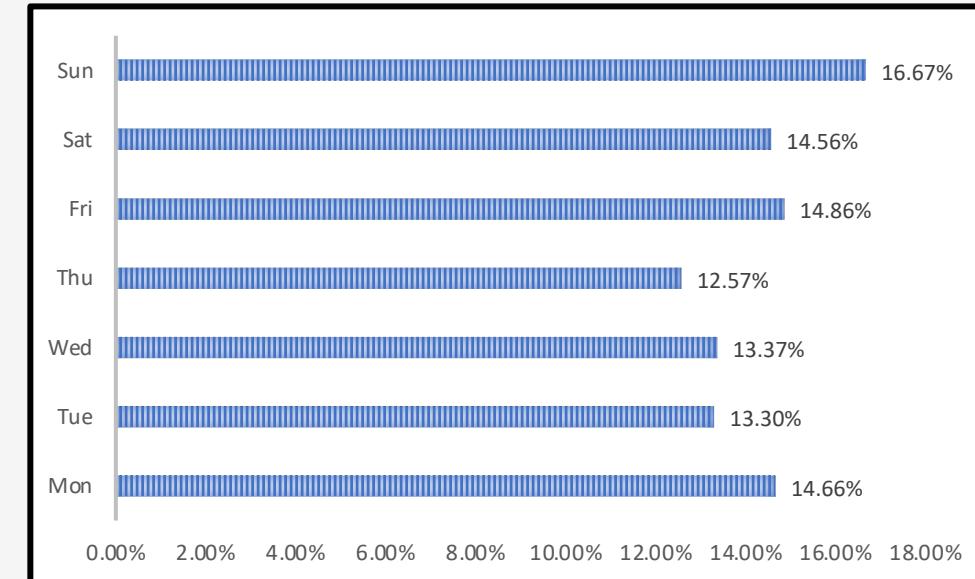
“Torture the data, and it will confess to anything”.

- Ronald Coase

Online Order flow as per the time of the day:



Online Order flow as per the days of the week:

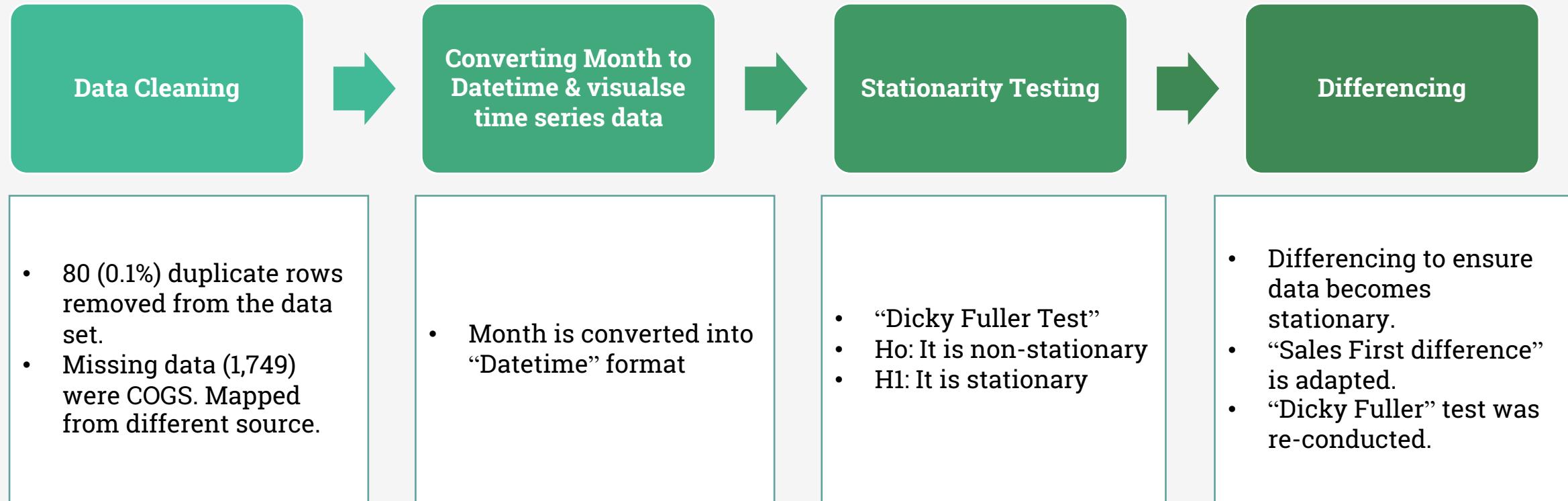


- **70% of the orders flow in from noon to midnight.**
- **10% of the orders come in during midnight to six in the morning.**
- It is all about the **weekend!**
- **NO Monday Blues!**

08. Data Preparation

“Give me six hours to chop down a tree and I will spend the first four sharpening the axe”.

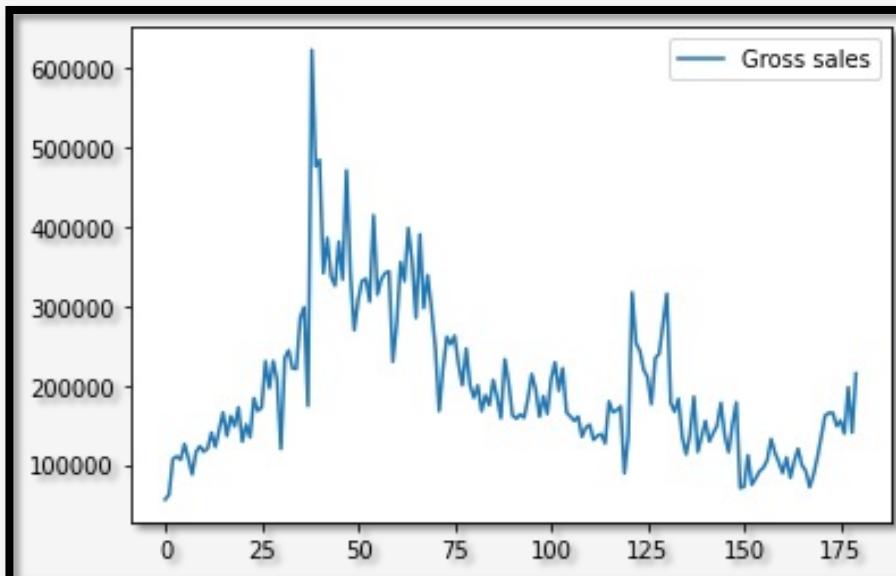
- Abraham Lincoln



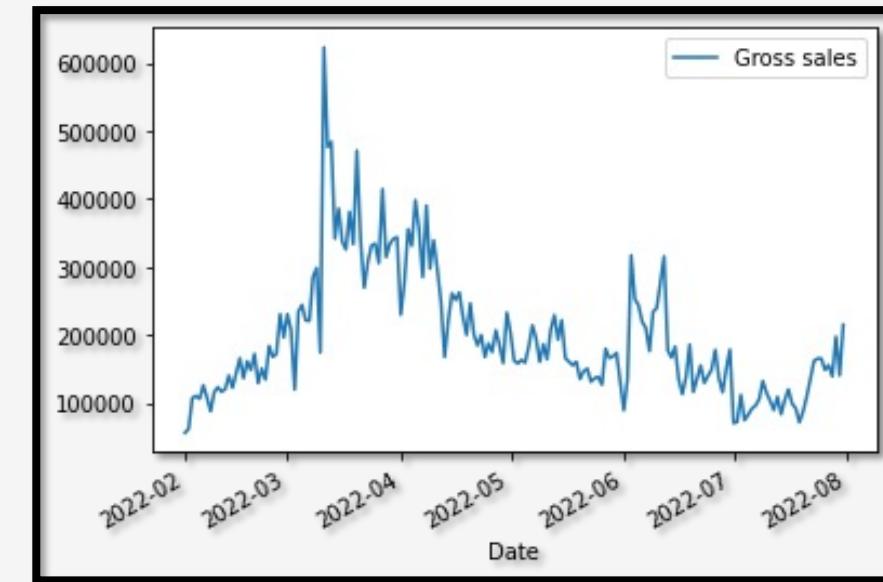
“No data is clean, but most is useful”.

- Dean Abbott

Converting Month to Datetime & visualise time series data:



Gross sales before converting to Datetime



Gross sales post converting to Datetime format

Stationarity Testing & Differencing:

“No data is clean, but most is useful”.

- Dean Abbott

Ho: It is non-stationary
H1: It is stationary



```
ADF Test Statistic : -1.9091341775456034
p-value : 0.3277763816273437
#Lags Used : 12
Number of Observations Used : 167
weak evidence against null hypothesis, time series has a unit root, indicating it is non-stationary
```

P-value is 32% which is greater than 5% hence the data is **non-stationary**.



Differencing



```
ADF Test Statistic : -3.4802686418808784
p-value : 0.00850853495400031
#Lags Used : 6
Number of Observations Used : 167
strong evidence against the null hypothesis(Ho), reject the null hypothesis. Data has no unit root and is stationary
```

P-value is 0.8% which is lesser than 5% hence the data is **stationary**

09. Modelling



ARIMA



SARIMAX

ARIMA (Auto Regressive Integrated Moving Average):

- predicts the target variable by utilizing its historical records.
- Auto-Regressive means that it utilizes “lag values” to forecast.
- MA component utilizes “lagged forecast errors” for its prediction.
- Integrated (I) combines both “AR” & “MA” components together.



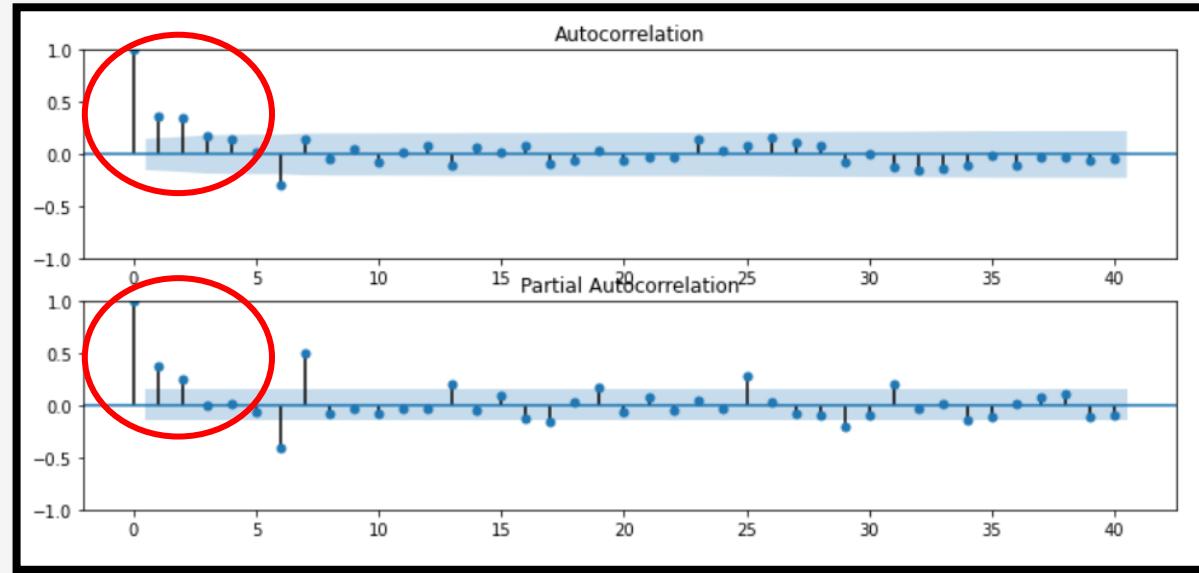
SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous factors):

- SARIMAX is an uplifted version of the ARIMA model.
- Sales data has a seasonal aspect
- Hence, SARIMAX would be the best approach.

10. Model Evaluation

Let's assess the models!

ARIMA:



ACF and PACF plots

- The order of autoregression in the PACF plot is two.
- This means the lags do not cross the threshold post the second lag.
- Model's "AR" component has fitted well.
- Order of autoregression in ACF is two as well.
- This translates that the model's "MA" component is fitting well too.

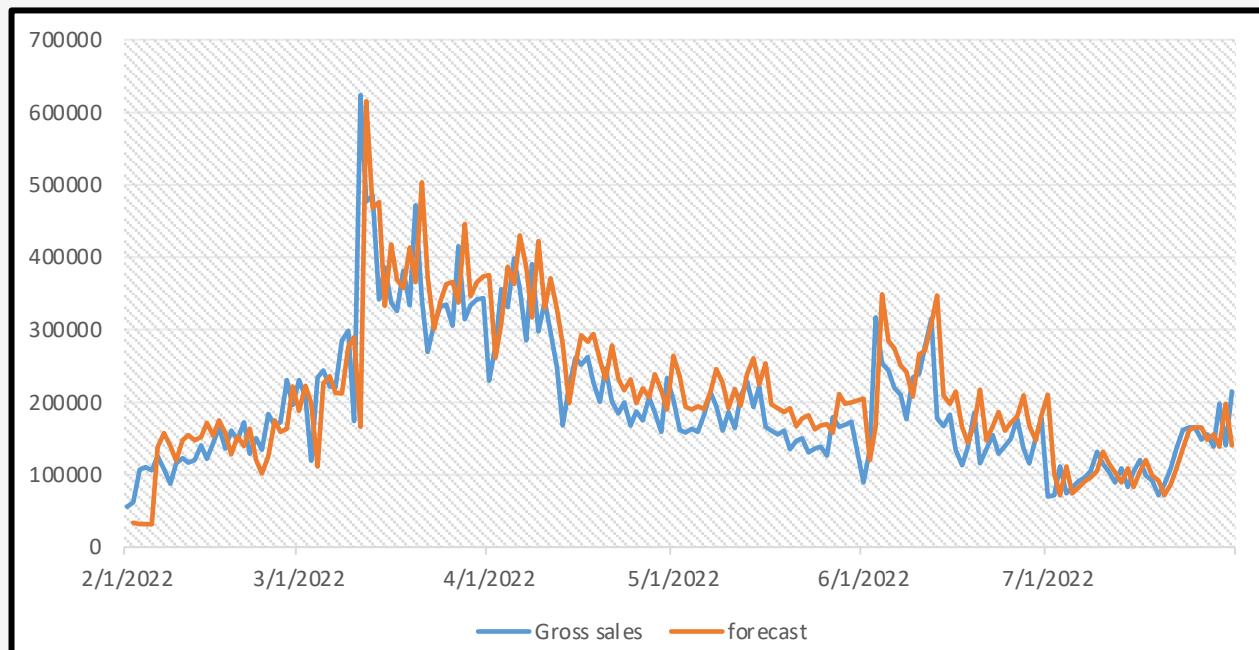
Let's assess the models!

SARIMAX:

SARIMAX Results									
Dep. Variable: Gross sales		No. Observations: 180							
Model: ARIMA(1, 1, 1)		Log Likelihood -2195.861							
Date: Tue, 23 Aug 2022		AIC 4397.722							
Time: 05:28:36		BIC 4407.284							
Sample: 0 - 180		HQIC 4401.599							
Covariance Type: opg									
	coef	std err	z	P> z 	[0.025 0.975]				
ar.L1	0.0846	0.080	1.060	0.289	-0.072 0.241				
ma.L1	-0.6246	0.079	-7.878	0.000	-0.780 -0.469				
sigma2	2.724e+09	6.9e-11	3.95e+19	0.000	2.72e+09 2.72e+09				
Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 3032.76									
Prob(Q): 0.95		Prob(JB): 0.00							
Heteroskedasticity (H): 0.35 Prob(H) (two-sided): 0.00									
Skew: 2.59 Kurtosis: 22.49									

SARIMAX results

- AR lag one is non-significant since the p-value is greater than 5%.
- MA lag one is significant since the p-value is lesser than 5%.
- “Ljung-Box” Statistical approach, probability score is 0.95.
- Hence it translates that the model has fitted well.



SARIMAX (Actual vs. Predicted Sales).

11. Analysis and Results

“We are surrounded by data but starved for insights”.

- Jay Baer

Category Insights:

Gross Sales

- Men's wear: **Stable** and the **biggest contributor** to sales.
- Women's wear: **Drastically dropped** since April.
- Kid's wear: Performed well in the **first 3 months** and **dipped since April**.
- Ethnic wear: A **slow start in the beginning** and **increase in top line** in May and June.
- Non-Apparel categories: **Contribution is <6%**.

Discounts

- Men's wear: **Steady share of business with minimum discounts.**
- Ethnic wear: Topline is achieved **via running discounts.**
- Other categories: Maintained a **healthy discount %.**

Net profit

- Ethnic wear: **Least profitable - high discounts** run in May and June.
- **Ethnic wear** is depleting the NP %
- Highest COGS are also ethnc wear.
- Other categories: **Healthy Net profit margin.**

“We are surrounded by data but starved for insights”.

- Jay Baer

Operational insights:

RTO & Returns

- Final sales impacted the top line is **63% of gross Sales**.
- **45%** of confirmed orders have returned during the past six months.
- Industry standard of **<25%**.

Mode of Payment

- COD is the **most preferred mode of payment**
- Room for **cancellation / change of mind or circumstances**.
- **Ease of payment** to the delivery partner via UPI payments.

Order Flow

- **70%** of the orders flow in from **noon to midnight**.
- **10%** of the orders come in during **midnight to six in the morning**.
- It is all about the **weekend!**
- **NO Monday Blues!**

12. Conclusions and Future Scope

“All good things come to an end”.

- Geoffrey Chaucer

Conclusion:

- **SARIMAX** is the best sales forecasting technique to adopt since it **smoothens out most of the volatility** in the data and most of all takes care of the **seasonality aspect**.
- Provided **meaningful** and **actionable insights** to achieve **better profitability**.
- ***Better forecasting will directly lead to better inventory management.***
- This will **influence a positive topline** as well as a **positive bottom line**.

Future scope and points of discussion:

- Make the model stronger by **adding MoM sales**.
- Expand the data set **to other online channels and offline stores**.
- Once the **analytical culture** is built within the organization, more forecasting techniques such as **Long Short-Term Memory (LSTM)**, **Linear Regression**, and **Random Forest** can be explored.

Bibliography | Webliography

Bibliography:

- Belgamwar, T. (2021). Inventory Management using Demand Sales Forecasting. In International Journal of Operations Management and Services (Vol. 11, Issue 1).
<http://www.ripublication.com>
- Bug, J. E. P. (2016). Application of predictive analytics to sales forecasting in the fashion business. <https://www.researchgate.net/publication/325100494>
- Deloitte Digital. (2022). Apparel Trends 2025. <https://www.deloittedigital.com/content/dam/deloittedigital/us/documents/blog/blog-20200610-apparel-trends.pdf>
- Fattah, J., Ezzine, L., Aman, Z., el Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10.
<https://doi.org/10.1177/1847979018808673>
- IBM. (n.d.). Cross-industry standard process for data mining Lifecycle. Retrieved August 18, 2022, from <https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview>
- McKinsey & Company. (2022). The State of Fashion 2022. <https://www.mckinsey.com/~/media/mckinsey/industries/retail/our%20insights/state%20of%20fashion/2022/the-state-of-fashion-2022.pdf>
- Shakti, S. P., Hassan, M. K., Zhenning, Y., Caytiles, R. D., & N.Ch.S.N, I. (2017). Annual Automobile Sales Prediction Using ARIMA Model. International Journal of Hybrid Information Technology, 10(6), 13–22. <https://doi.org/10.14257/ijhit.2017.10.6.02>
- Statista Digital Market Outlook. (2021). Fashion eCommerce report 2021. <https://www.statista.com/study/38340/e-commerce-report-fashion/>
- Tony Yiu. (2020, April 26). Understanding ARIMA (Time Series Modeling). <https://towardsdatascience.com/understanding-arima-time-series-modeling-d99cd11be3f8>
- Wazir Advisors. (2022). Wazir Report - The Road to 2025. 1–32. <https://wazir.in/pdf/Wazir%20Report%20-%20The%20Road%20to%202025.pdf>
- YUGESH VERMA. (2021, July 30). Complete Guide to SARIMAX in Python for Time Series Modeling. <https://analyticsindiamag.com/complete-guide-to-sarimax-in-python-for-time-series-modeling/>

Sales Analytics to drive Profitability - A case study of a Fashion E-Commerce Retailer

ORIGINALITY REPORT

6%	4%	2%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	global.oup.com Internet Source	1%
2	www.statista.com Internet Source	1%
3	Submitted to University of Aberdeen Student Paper	1%
4	Jamal Fattah, Latifa Ezzine, Zineb Aman, Haj El Moussami, Abdeslam Lachhab. "Forecasting of demand using ARIMA model", International Journal of Engineering Business Management, 2018 Publication	1%
5	Submitted to University of College Cork Student Paper	<1%
6	www.science.gov Internet Source	<1%
7	Submitted to Higher Education Commission Pakistan Student Paper	<1%

8	Submitted to University of East London Student Paper	<1 %
9	www.miljo.fi Internet Source	<1 %
10	kmpls.vse.cz Internet Source	<1 %
11	mdpi-res.com Internet Source	<1 %
12	store.ectap.ro Internet Source	<1 %
13	www.sersc.org Internet Source	<1 %
14	Submitted to SP Jain School of Global Management Student Paper	<1 %
15	www.projectpro.io Internet Source	<1 %

Exclude quotes On Exclude matches < 10 words
 Exclude bibliography On

Additional Information | Plagiarism score



*Thank
you!*