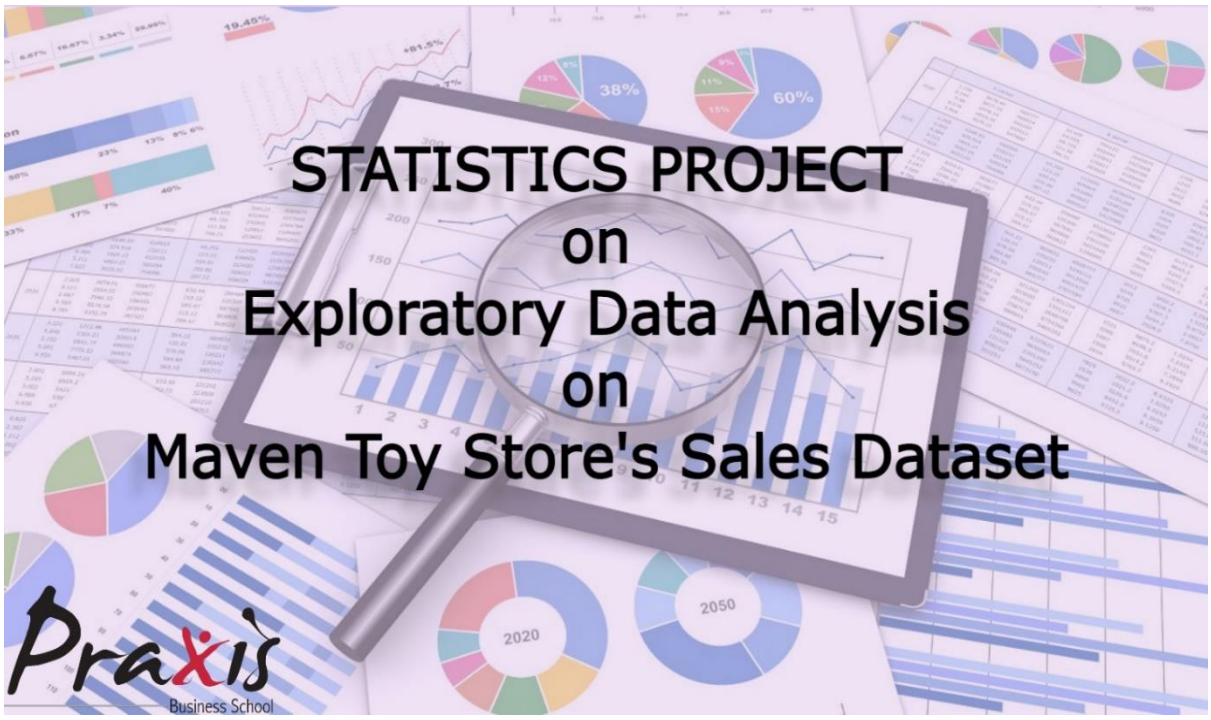
# Praxis Business School

## PGPDS - Spring Kolkata 2022




Submitted by

**Agniva Roy - C22001**

**Preetam Saha – C22018**

[Guided by Dr. Sayantani Roy Chowdhury]

## Goal Of the Project:

There has been rapid growth in technology within the last few decades. This has affected the process of storing data a lot. But the huge amounts of stored data are of no value if we cannot draw valuable insights from them. In this project, we intend to explore a dataset and detect and uncover the underlying structure of the dataset. So, in a nut shell, as data scientist, using EDA we can understand the data in a detailed way and take some actionable decisions for the overall growth of the organization.

In this dataset, we have got the sales of a fictional toy store franchise. We want to use EDA to see how the sales and profits are varying and getting affected based on various parameters like product category, product group, location of store etc. We also want to see the small percentage of products that are responsible for the majority of the sales. Using this we want to get valuable insights that would help the franchise to increase their profits.

## About Maven Analytics:


Maven Analytics is an E-Learning company based in Boston, Massachusetts. Since 2014 they have helped students and teams across 150+ countries develop the most sought-after analytics and business intelligence skills, through on-demand courses, skills assessments, curated learning paths, and enterprise training. Maven Analytics has been featured by USA TODAY as one of the top 10 online education companies revolutionizing the EdTech industry.

Link to the original article-https://www.usatoday.com/story/sponsor-story/level-up/2021/12/09/10-online-education-companies-revolutionizing-industry/6431032001/

## Data Collection:

There are mainly two types of data collection namely primary and secondary. Here we are using fictional Maven's Toy store's data that is collected from Maven Analytics Toy Challenge, which is a secondary data source.

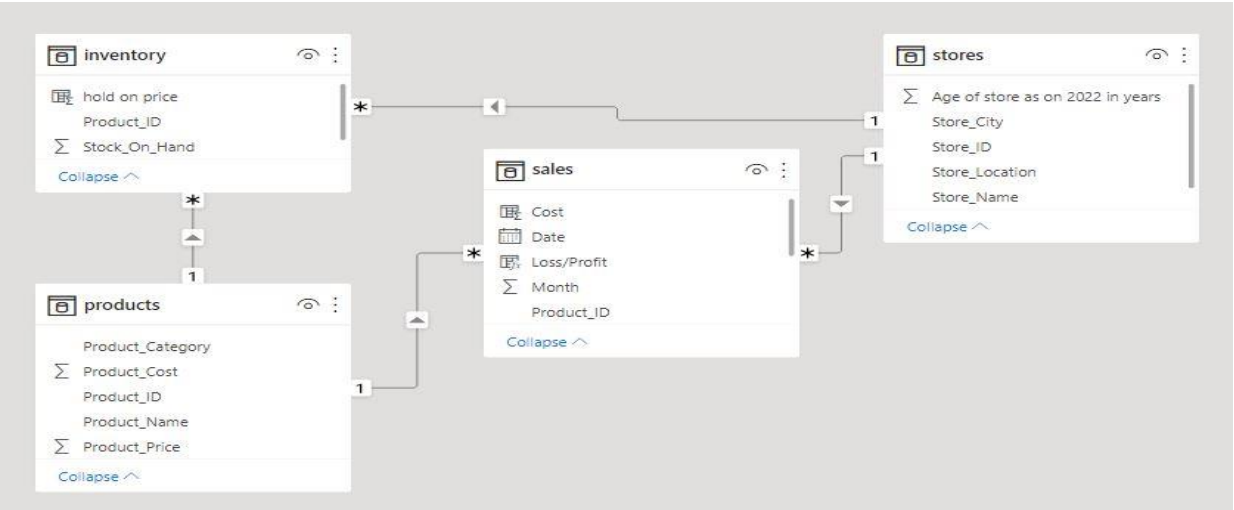
# About the dataset:


The dataset contains the sales & inventory data from Maven Toys, a fictitious chain of small toy stores in Mexico. This dataset was used in the Maven Toys Challenge.

This dataset contains 4 tables, in CSV format:

- The Products table contains the 35 products sold at Maven Toys (each record represents one product), with fields containing details about the product category, cost, and retail price
- The Stores table contains the 50 Maven Toys store locations (each record represents one store), with fields containing details about the store location, type, and date it opened
- The Sales table contains the units sold in over 800,000 sales transactions from January 2017 to October 2018 (each record represents the purchase of a specific product at a specific store on a specific date)
- The Inventory table contains over 1,500 records that represent the stock on hand of each product in each store.

These 4 tables are interconnected with each other via relationships. The following relational schema shows the relationships the tables have with each other.



## Tool used:

In this project, we have used Microsoft Power Bi Desktop for the analysing and visualizing the dataset

Power BI Desktop

# Exploratory Analysis of Data:

## Scorecard showing brief summary:


In this period of time, almost **14 million** dollars' worth of sales have been made.

Profits around **4 million** dollars have been generated.

Around **1091k** units have been sold. These involved around **829k** transactions.

## Descriptive Statistics of Sales and Profits per Transaction:

## Product Categories

| Product_Category | Avg | Max | Min | Std |
|---|---|---|---|---|
| Toys | $23.02 | $879.78 | $5.99 | $21.10 |
| Electronics | $22.69 | $155.94 | $14.99 | $12.83 |
| Sports & Outdoors | $16.54 | $139.93 | $8.99 | $9.73 |
| Games | $14.18 | $499.75 | $6.99 | $12.09 |
| Art & Crafts | $12.26 | $149.94 | $2.99 | $12.10 |

| Product_Category | Avg | Max | Min | Std |
|---|---|---|---|---|
| Toys | $4.88 | $138 | $1 | $3.55 |
| Sports & Outdoors | $3.85 | $64 | $1 | $3.15 |
| Games | $4.29 | $150 | $2 | $4.03 |
| Electronics | $10.11 | $72 | $5 | $6.57 |
| Art & Crafts | $3.41 | $60 | $1 | $3.24 |

We can see that the toys product category has the highest average sale value per transaction while Games category has the lowest average sale value per transaction. The standard deviation for toys product category is also the most, which suggests that transactions that involve toys have a wide variability ranging from cheap items to very expensive items. While the deviation is least for the Sports and Outdoors category, which suggests that compared to the other categories it has less variability. The average profit per transaction is highest in the electronics category while it is the lowest in the Art & Crafts category.

## Product Groups

| Product_group | Avg | Max | Min | Std |
|---|---|---|---|---|
| Cheap | $9.50 | $329.7 | $2.99 | $7.93 |
| Expensive | $43.44 | $879.78 | $24.99 | $27.13 |
| Moderate Price | $20.83 | $499.75 | $12.99 | $11.75 |

| Product_group | Avg | Max | Min | Std |
|---|---|---|---|---|
| Cheap | $3.13 | $150 | $1 | $3.18 |
| Expensive | $6.33 | $110 | $4 | $3.91 |
| Moderate Price | $6.31 | $150 | $2 | $5.04 |

As it can be expected, expensive products have the highest average sale value per transaction while it is lowest for cheap products. The variability is also the most in expensive products and the least in cheap products. But we see that expensive products and moderate price products have the almost same average profit value per transaction.

## Analysis by Product Categories:

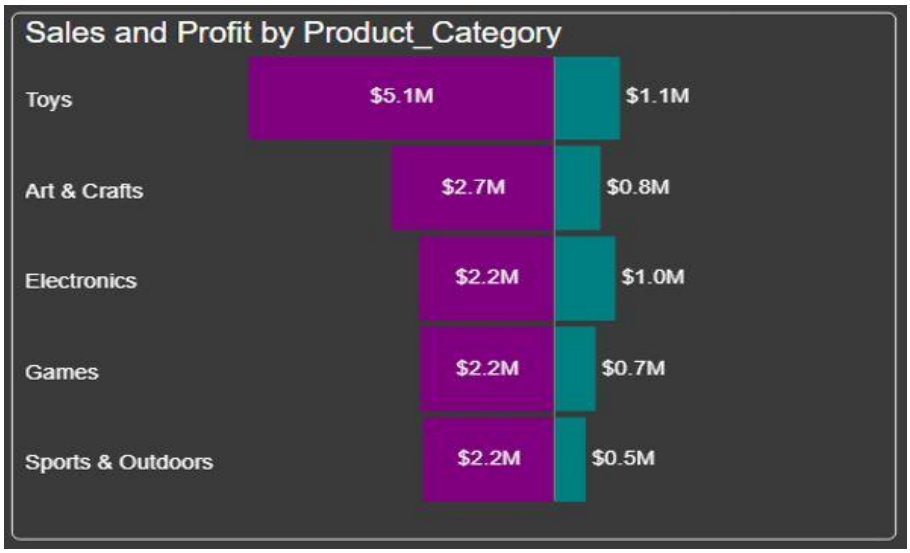
## 1.Sales:


*Tree map showing the sales of each product category*

From this tree map, we can see that

- The **Toys** category has the most sales
- **Art & Crafts** come at the second
- **Electronics**, **Games** and **Sports & Outdoors** have almost same sales
- The sales of toys are more than the sales of **Electronics** and **Games** categories combined.

## 2. Combined Sales and Profit:

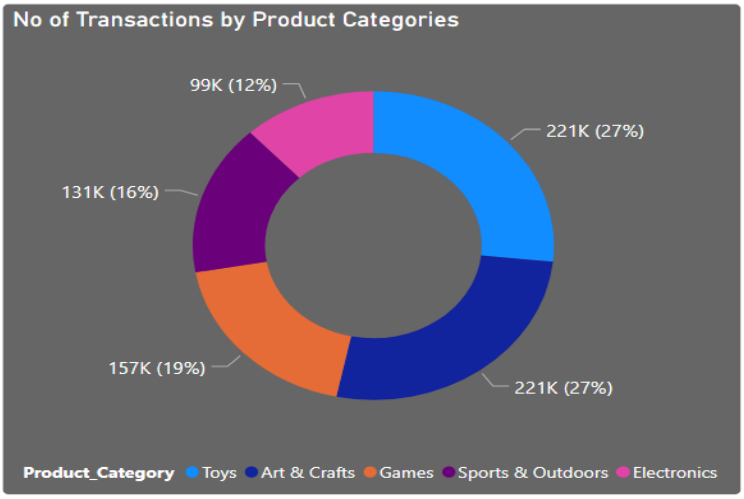



*Tornado chart showing the sales and profits of each product category*

From the tornado chart, we can see that

- **Toys** has the most sales as well as the most profits
- Although **Art & Crafts** is the 2nd in terms of sales, the **Electronics** category has more profit than the Art and Craft

## 3. Number of Transactions:




From this pie chart we can see that out of the total number of transactions that happened at various stores, the number of transactions involving **electronics** products were the lowest that is around **12%** of the total number of transactions. While it is highest for the **Toys** and **Art & Crafts** Product categories. They each account for around **27%** of the total transactions.

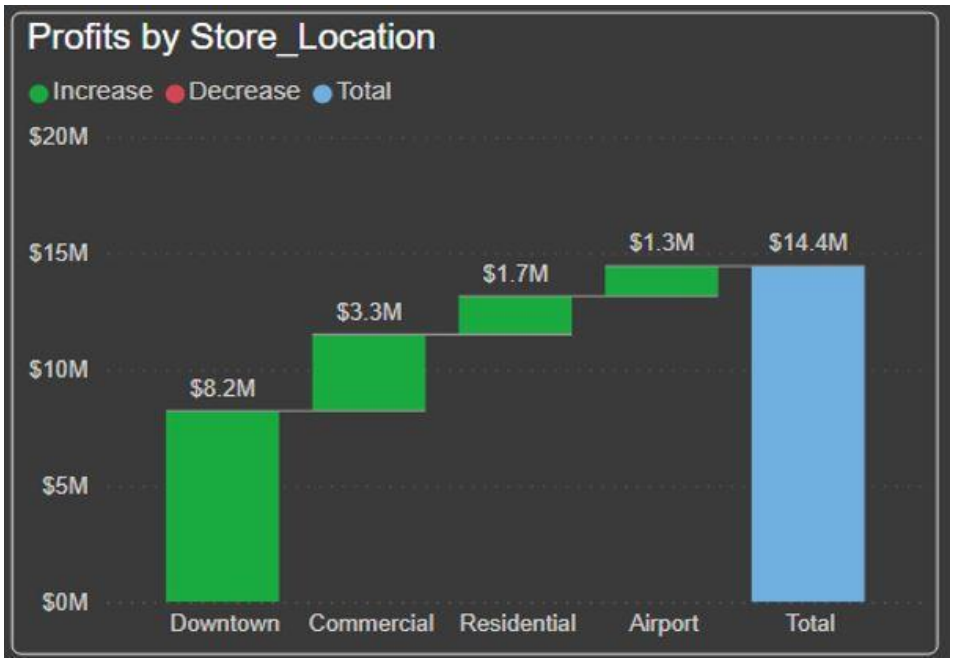
## Analysis by Store Locations:

## 1. Sales:


From this bar graph, we can see that

- Stores located in **downtown** areas have the most sales
- Stores located near **airports** have the least sales

## 2.Profit:




*Waterfall chart showing profits by different product categories*

From this chart, we have seen that downtown areas have generated the highest profits, that is around 8.2 million dollars.

The total profits generated is around 14 million dollars.

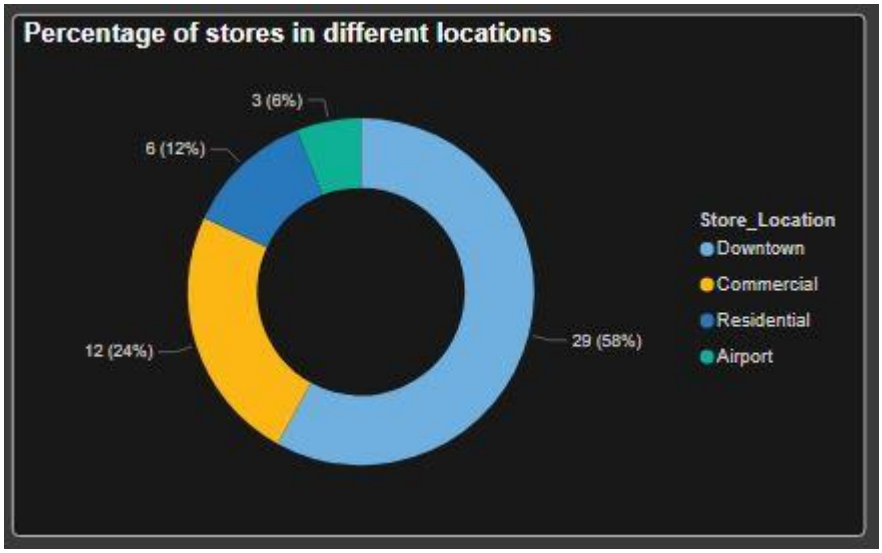
## 3.Percentage Contribution of Net Profit:


*Pie Chart showing Percentage Contribution of each store location towards profit*

From this pie chart, we can see that **56.02%** of the total profits is generated by stores located in **Downtown** areas.

Stores located near **airports** contribute to just **9.42%** of the total profits generated.

## 4.Percentage of Stores located in various Store Locations:




From the pie chart, we can see that an overwhelming majority of the stores are located in **downtown** areas that is **58%** of the total stores.

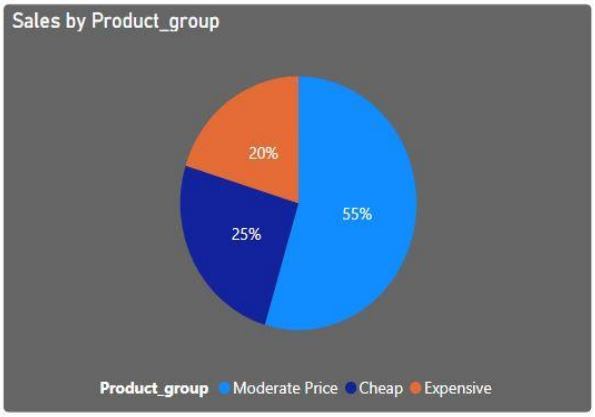
Areas near **airports** have the least number of stores which is just **6%** of the total stores.

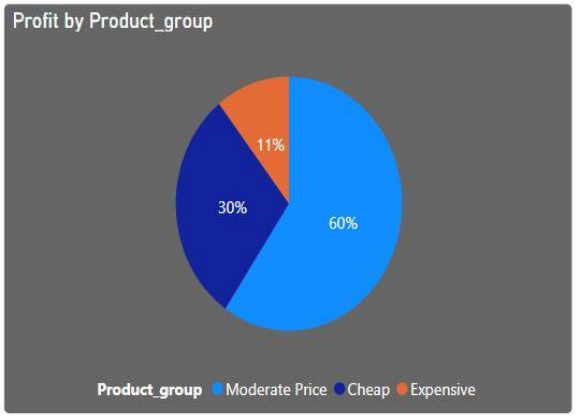
## Analysis by Product Groups:

Here, we have divided the 35 available products into 3 groups as per their selling prices

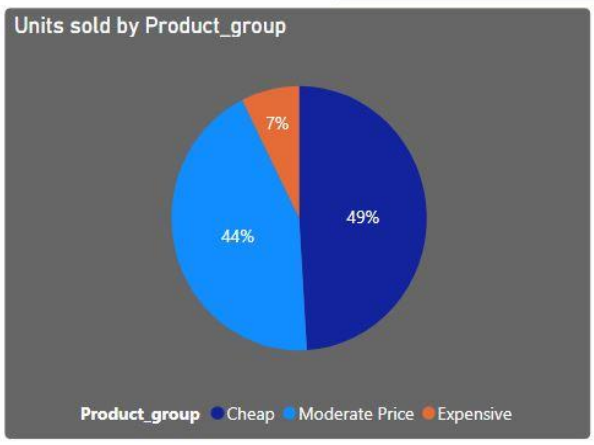
- If the price of the product is less than or equal to **11.99 dollars**, it is a **cheap product**
- If the price of the product is greater than 11.99 dollars and less than or equal to **23.99 dollars**, it is a product of **moderate price**
- If the price of the product is greater than 23.99 dollars, then it is **expensive product**

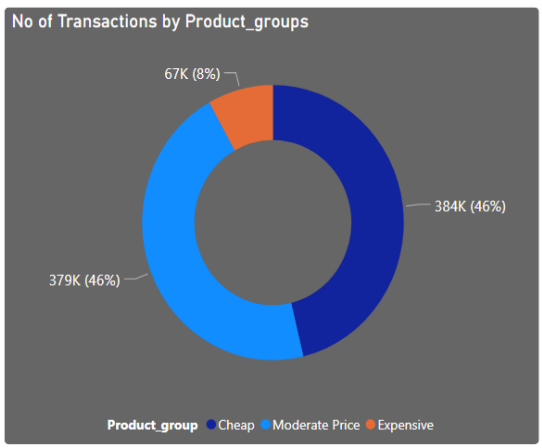
## Analysis of Sales, Profits and Units sold by Product Groups:




As we can see from pie charts, majority of the sales and profits came from products which were moderately priced-which is around **55%** of the total sales and **60%** of the total profits respectively.

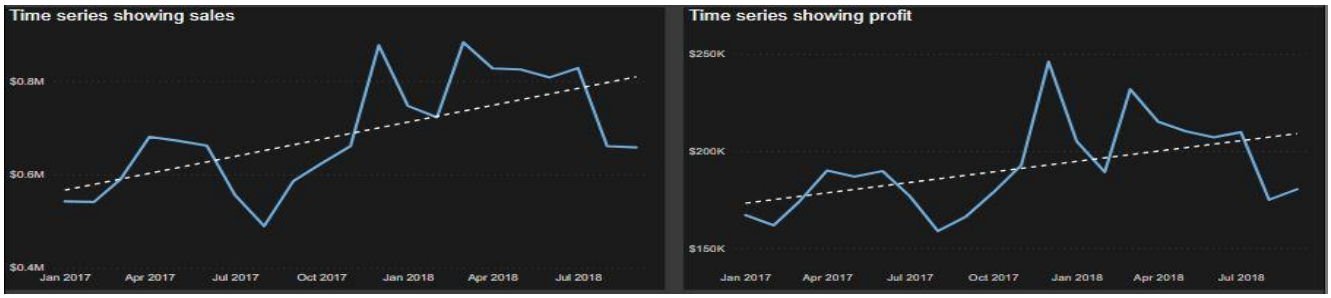



From this pie chart we can see that expensive products were the least sold in terms of units, that is just **7%** of the total units sold. Units sold were maximum in case of cheap products (around **49%** of the total units sold). This is generally expected.

we can also see that very less percent of total transactions involves expensive products. It is around just **8%**. On the other hand, transactions involving cheap and moderate priced products contribute for around **46%** each, of the total number of transactions.

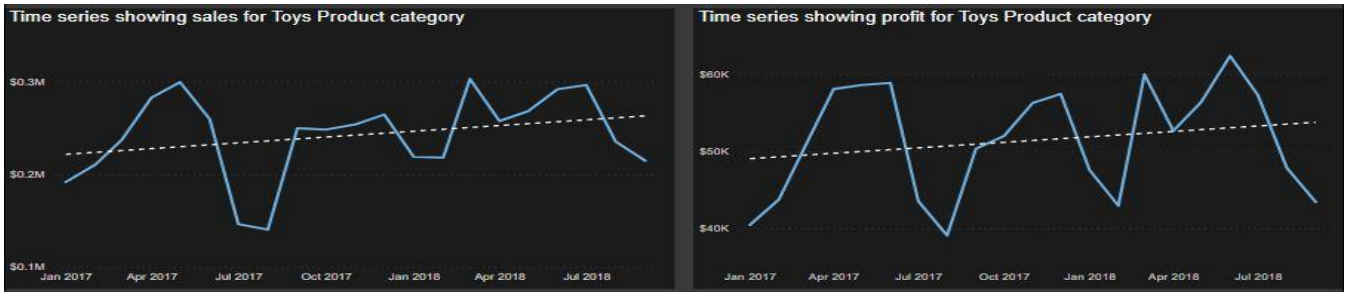
## Time Series Analysis of the Sales and Profits:




*Time Series showing total sales and total profits*

From the time series graph, we can see that on an average, the sales and profits have increased over time.

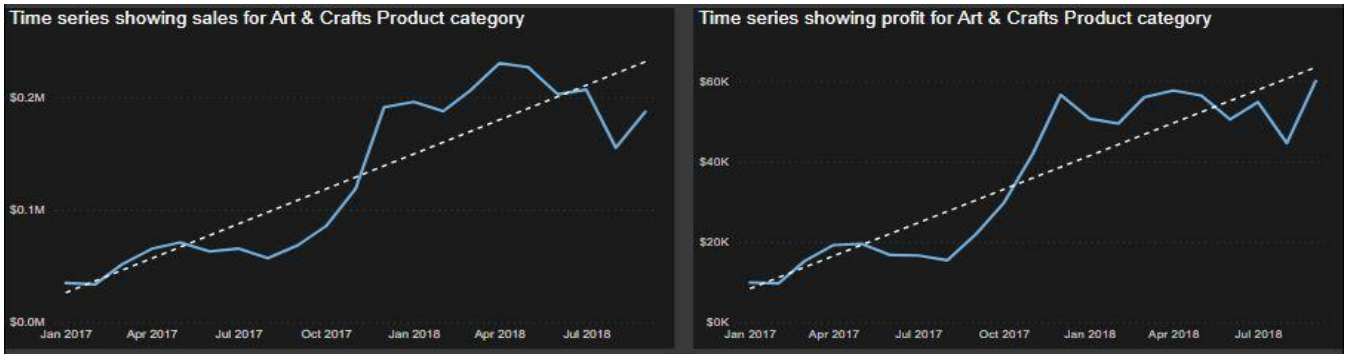
## Toys:




*Time Series showing sales and profits for Toys*

From the time series graph, we can see that on an average, the sales and profits of **toys** category have increased over time.

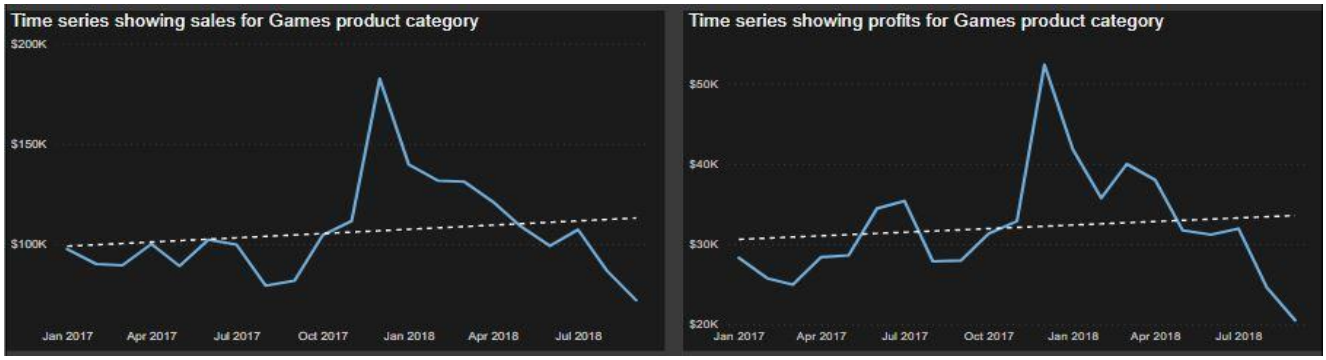
## Art & Crafts:




*Time Series showing sales and profits for Art & Crafts*

From the time series graph, we can see that on an average, the sales and profits of **art & crafts** category have increased over time.

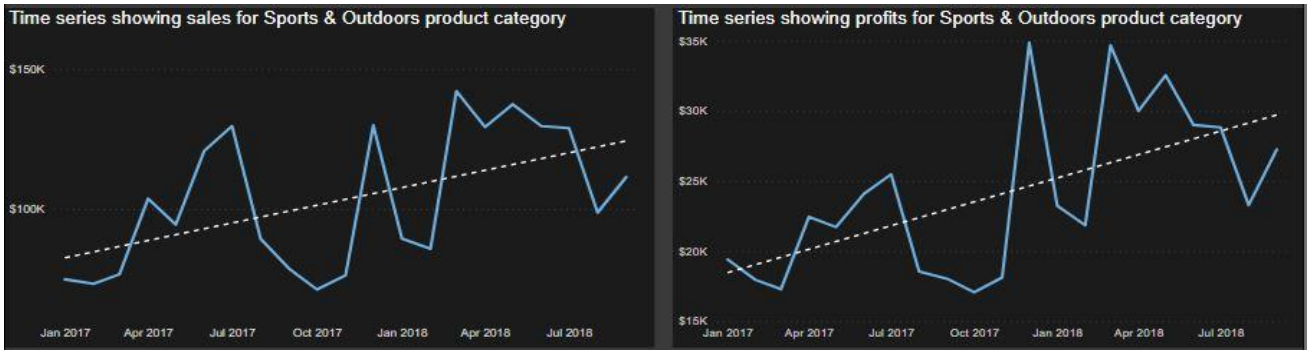
## Games:




*Time Series showing sales and profits for Games*

From the time series graph, we can see that on an average, the sales and profits of **games** category have almost remained constant

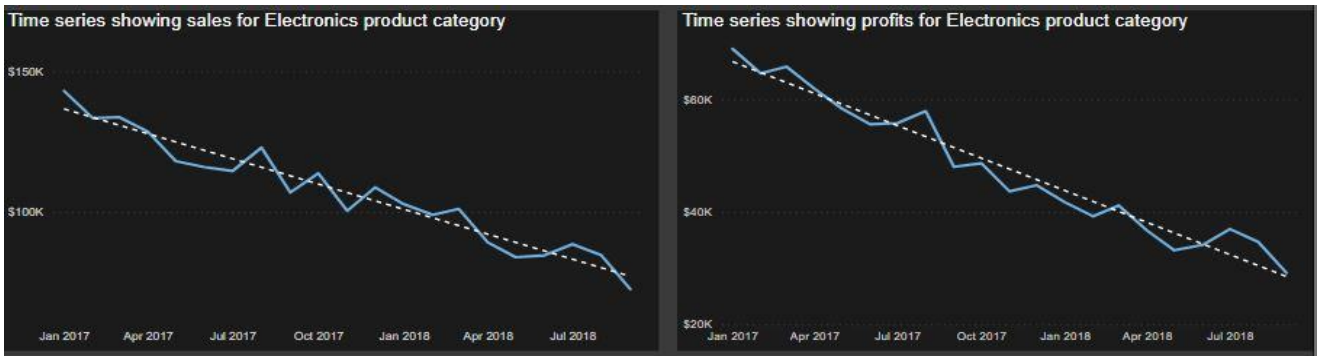
## Sports and Outdoors:




*Time Series showing sales and profits for Sports & Outdoors*

From the time series graph, we can see that on an average, the sales and profits of **sports and outdoors** category have increased over time.

# Electronics:

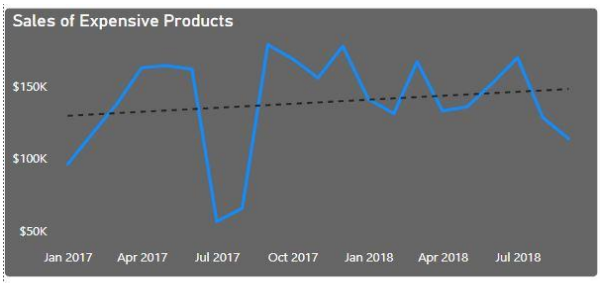


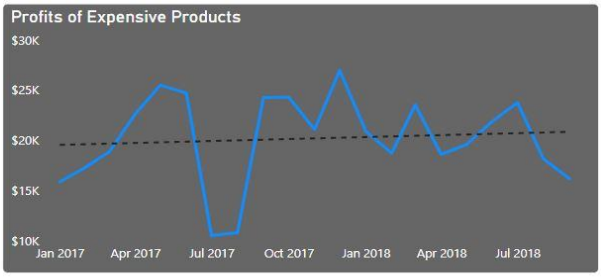

*Time Series showing sales and profits for Electronics*

From the time series graph, we can see that on an average, the sales and profits of **electronics** category have decreased over time.

# Time Series Analysis of the sales and profits of each product group:

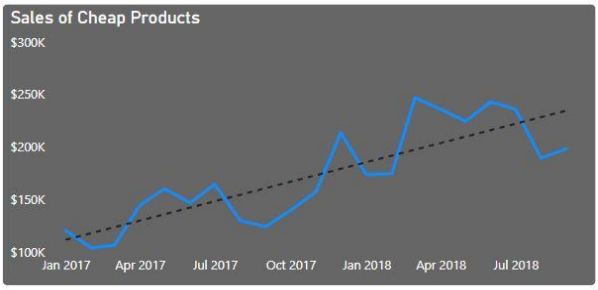
## Expensive products:

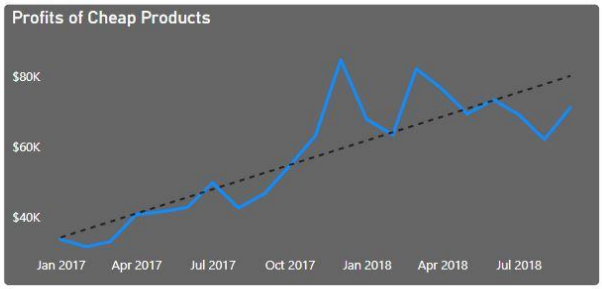



The sales and profits have increased over time but it is not that great

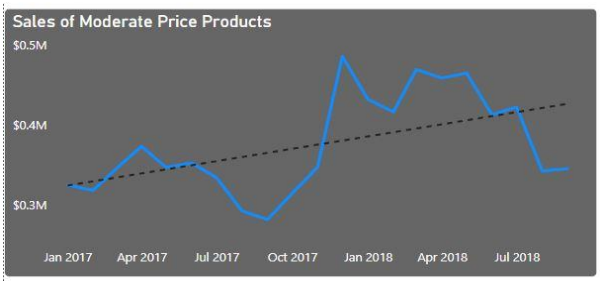
## Cheap products:

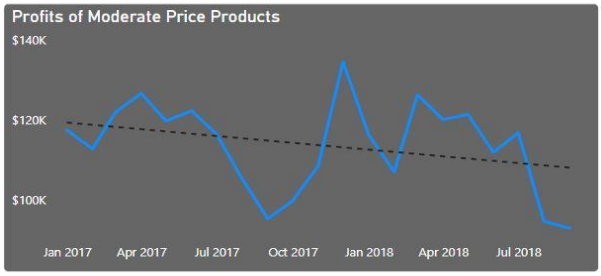



Sales have increased sharply

## Moderate price products:

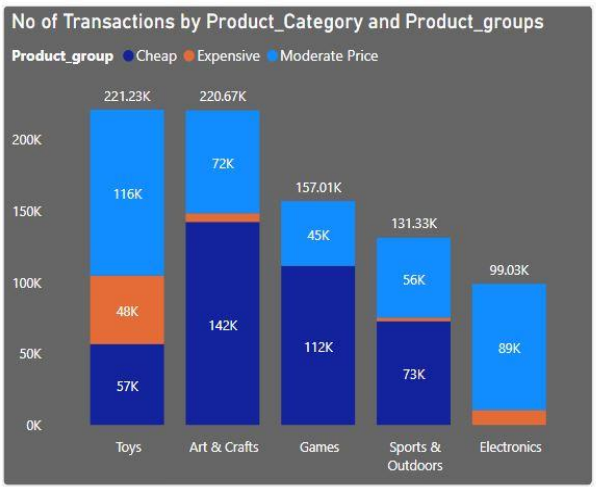


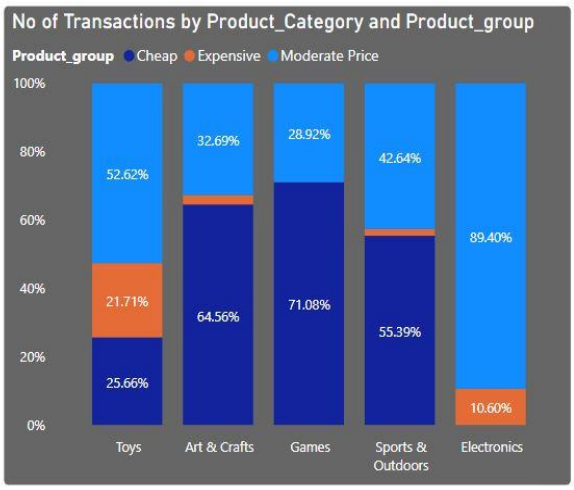




Sales have increased, but the profits have decreased.

# Analysis by Product Categories and Product Groups:

## 1.No of Transactions:



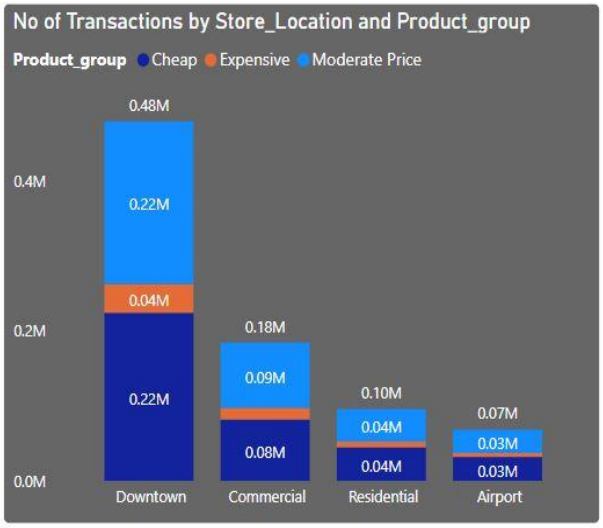



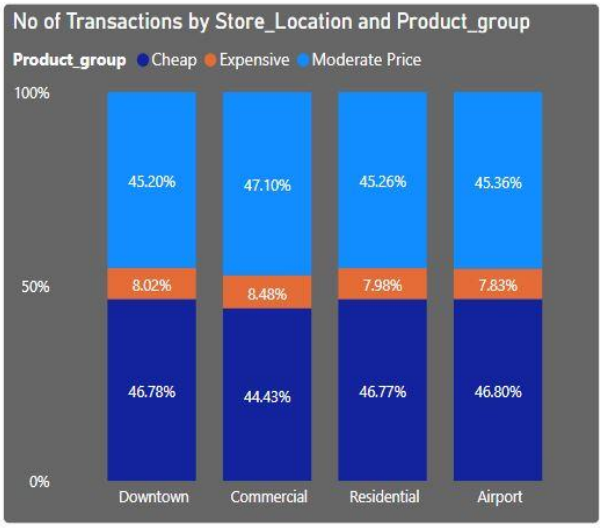

We can see that the number of transactions in **toys** and **arts & crafts** are almost similar that is close to **220k**, while the **electronics** product category has the least number of transactions. Customers tend to buy toys and arts & crafts related products more frequently. Electronics products are brought less frequently compared to other products. **Games** have no expensive products while **Electronics** have no cheap products.

From the stacked bar chart and 100% stacked bar chart we can see that in **toys** and **electronics** categories**, moderately** priced items are the most frequently brought while in other product categories **cheap** items are the most frequently brought. As expected, **expensive** items aren't brought that frequently. Even in the electronics category, the transactions involving expensive products is just around **10%**.

# Analysis by Store Location and Product Groups:

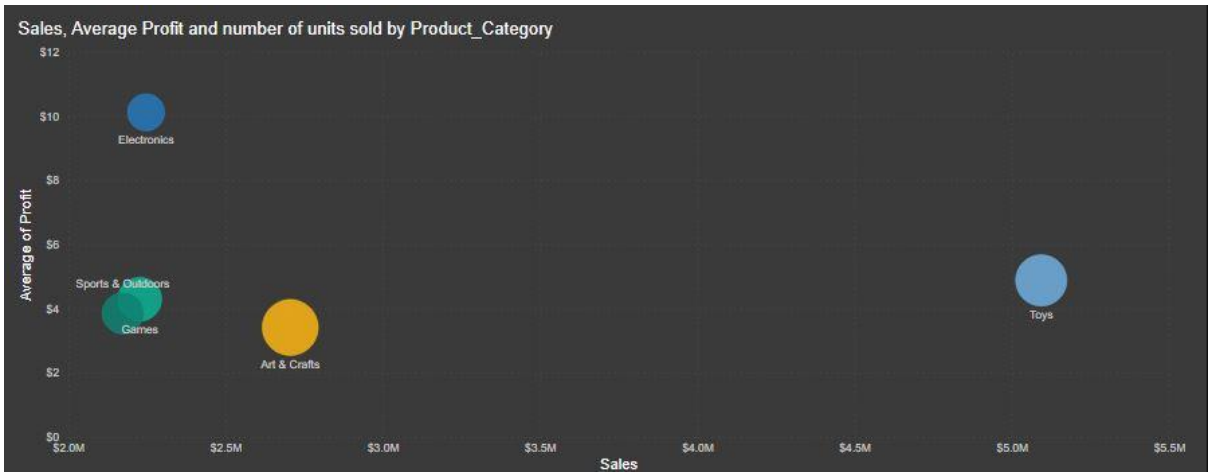
## 1.No of Transactions:



In store locations we can see the greatest number of transactions in **downtown areas** and the least in **airport areas**. However, there is almost no difference in the buying patterns. Cheap and moderately priced products are purchased more frequently than expensive products, and there seems no significant difference in the frequencies of their transactions in each of the location.

## Bubble chart of Sales vs Average Profit by Product Category:

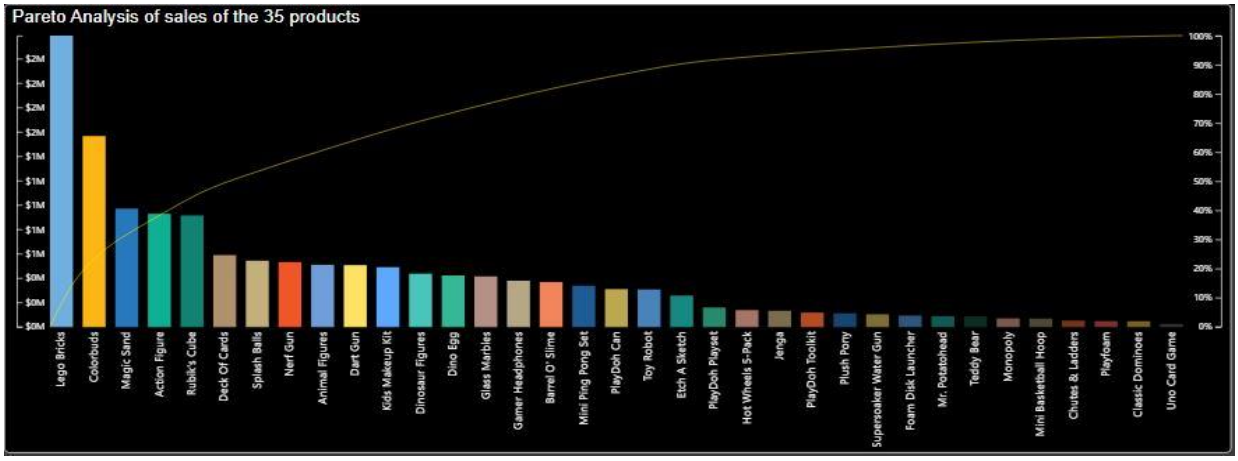



*Bubble chart showing Total sales vs average profit by product categories*

From this scatterplot, we can see that toys category is having the highest number of sales but its average profit is not that high. In comparison, electronics is not having high sales but its average profit is the highest. The size of the bubbles shows the number of units sold. Here the toys and art & crafts have a relatively big size. However, electronics has relatively less size, signifying that it the number of units of its products sold is not that high.

## Pareto Analysis:

The Pareto principle states that for many outcomes, roughly 80% of consequences come from 20% of causes. In other words, a small percentage of causes have an outsized effect. Here we are using Pareto principle to determine which are those small numbers of products out of so many available products that are responsible for the majority of the sales.

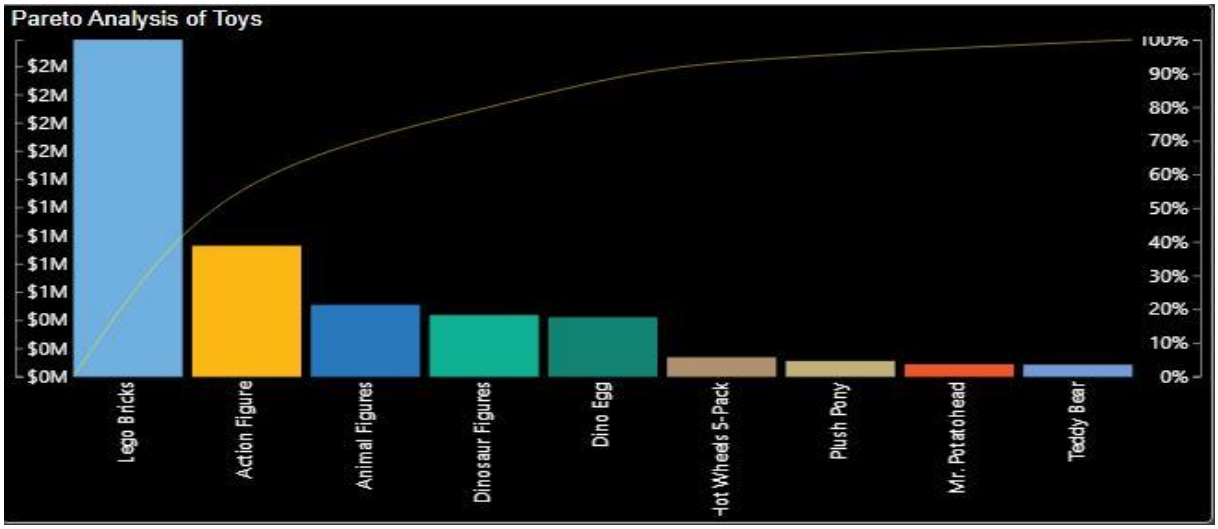



*Pareto Analysis of all the 35 available products*

Out of all the products**, Lego-bricks**, **Color buds**, **Magic Sand**, **Action Figure**, **Rubik's Cube** and **Deck of Cards** together contribute for around 50% of the total sales. Out of the 35 available products, these products are responsible for around 50% of the total sales.

## Pareto Analysis of Products by Product Categories:

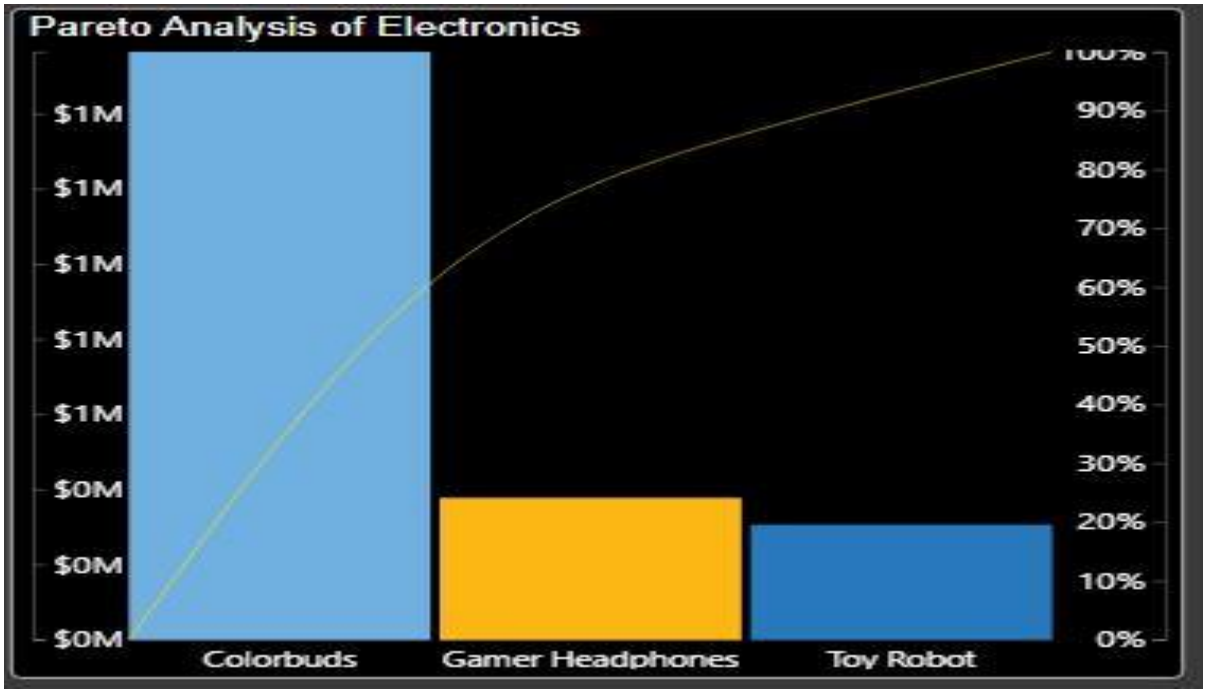
## Toys:




*Pareto analysis of products in the Toys category*

We can see from this Pareto chart that in the Toys product category, out of the **9** products available, the products **Lego Bricks** and **Action Figure** collectively contribute for around 65% of the total sales from toys.

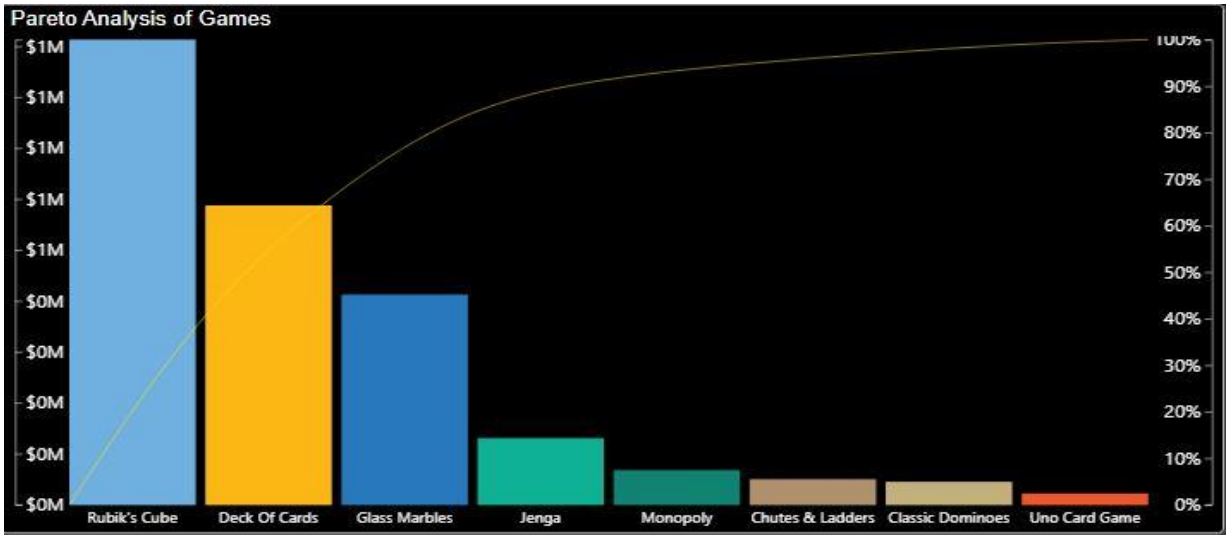
## Electronics:




*Pareto analysis of products in the Electronics category*

We can see from this Pareto chart that in the Electronics product category, out of the **3** products available, the product **Colorbuds** contribute for around 70% of the total sales from electronics.

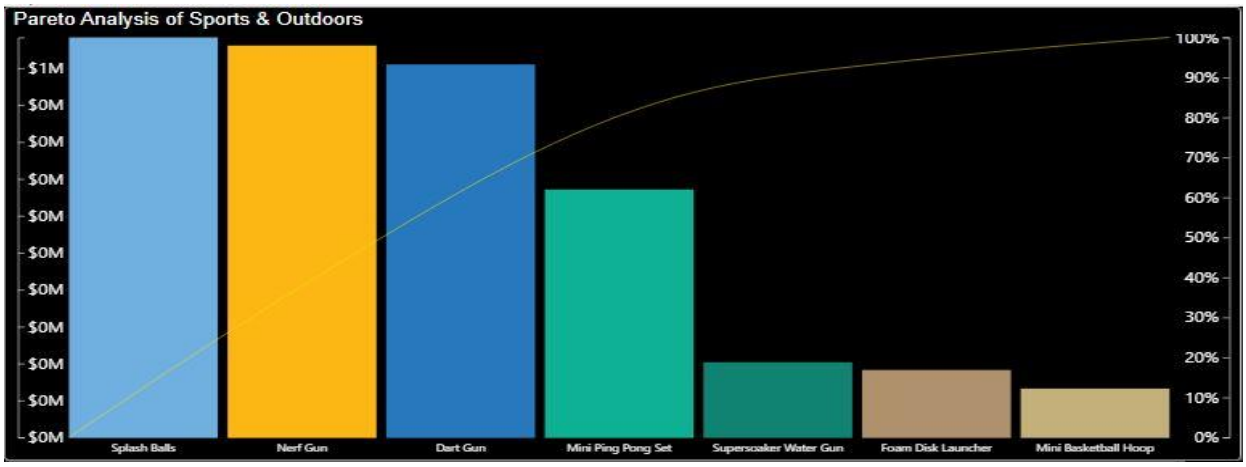
## Games:




*Pareto analysis of products in the Games category*

We can see from this Pareto chart that in the Games product category, out of the **8** products available, the products **Rubik's Cube** and **Deck of Cards** collectively contribute for around 67.38% of the total sales from games.

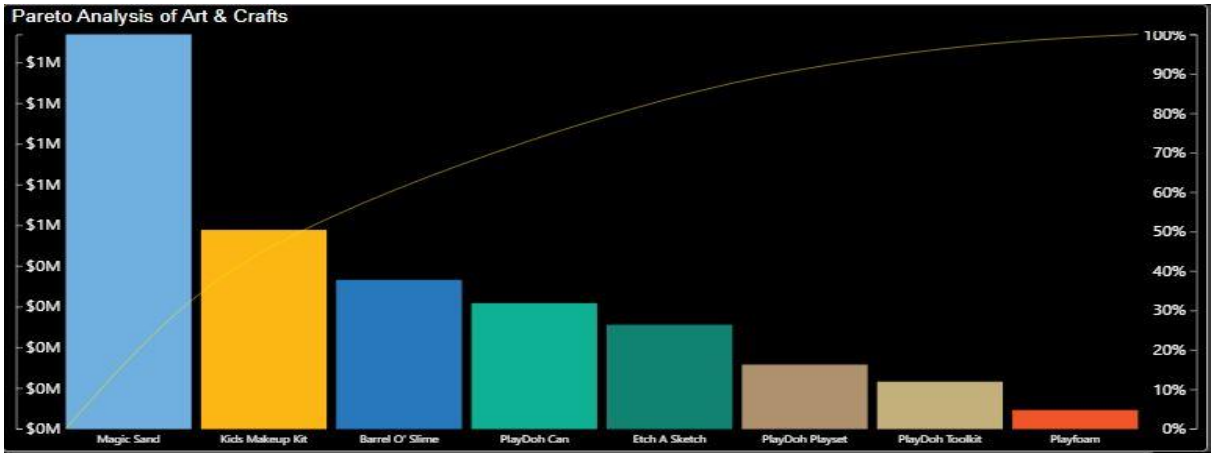
## Sports & Outdoors:




*Pareto analysis of products in the Sports & Outdoors category*

We can see from this Pareto chart that in the Games product category, out of the **7** products available, the products **Splash Balls, Nerf Gun** and **Dart Gun** collectively contribute for around 72.61% of the total sales from Sports and Outdoors.

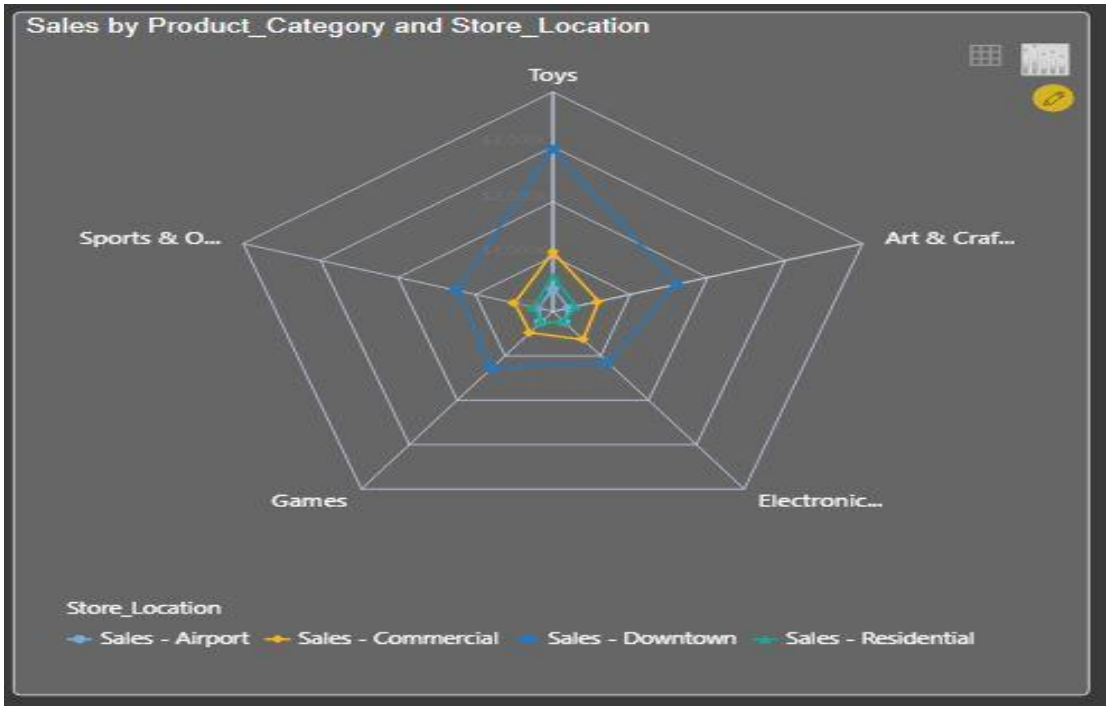
## Art & Crafts:




*Pareto analysis of products in the Art and Crafts category*

We can see from this Pareto chart that in the Games product category, out of the **8** products available, the products **Magic Sand** and **Kids Makeup Kit** collectively contribute for around 53.87% of the total sales from Art & Crafts.

## Radar Chart for sales of different Product Categories by Store Location:




*Radar chart showing the sales in different product categories by various store locations*

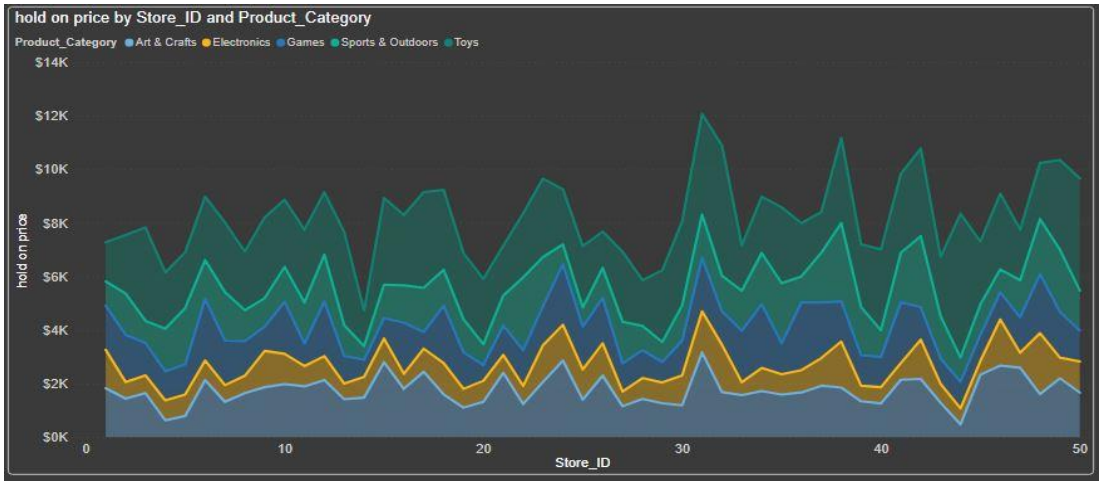
From this radar chart, we can see that sizes of the Sales Polygons for the Residential and Airport Areas are relatively smaller compared to the sales polygons for the Downtown and Commercial Areas. Regarding the categories, we can see that the radius corresponding to the Toys category is relatively more compared to the other categories.

## Radar Chart for Profit of different Product Categories by store location:


*Radar chart showing the profits in different product categories by various store locations*

From this radar chart, we can see that the size of the profit polygons corresponding to the profits in the residential and airport areas are relatively small compared to the profit polygons corresponding to the other 2 areas. We can also see that the radius corresponding to the profits related to Sports and Outdoor category is relatively smaller compared to the radii corresponding to the other categories.

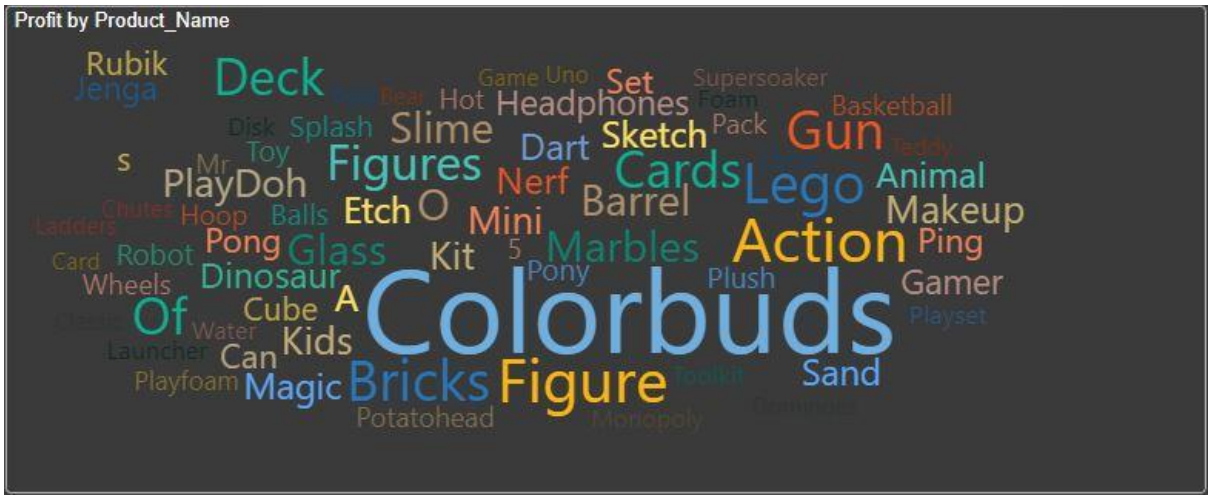
## Area chart showing hold on price in various stores by product category:




*Stacked area chart showing hold on price in various stores by product category*

From this area chart, we can conclude that the electronics category has the least worth of unsold products.

## Profits by Product Name:




*Word cloud representation of products by their profits*

We can see here the Colorbuds appears the biggest in this word cloud. This is because Colorbuds generates the most profits out of all other products

## Conclusions:

We can see that EDA gives us valuable insights that are not visible from a bird's eye view. From this type of analysis and more future analysis, businesses can take important decisions like **allocating budget to certain type of products, deciding the price band, deciding the areas to open new stores, deciding to add a new product and so on**.

Here are few suggestions that can be given

- Electronics seems to be profitable category. It does not rank that high in sales, it does not involve that many transactions, even most of its transactions involve moderately priced products, and still it is making considerable profits. They can probably add more products. It is a possible cash cow.
- At the same time, company should see why the sales and profits of electronics category has decreased despite doing so well in profits.
- Downtown areas seem to attract many sales. Company can put more focus on them. At the same time, they can identify why their sales are not that high in airport and residential areas and open more stores there.
- Company must decide on the pricing of moderately priced products as even though their sales have increased, their profits have decreased. It is possible that the COGS of moderately priced products is increasing.


THANK
YOU!