Name: Aqsa Younus

ERP: 18571

**Dataset:**

Pakistan's ECommerce Dataset:<https://opendata.com.pk/dataset/pakistan-largest-ecommerce-dataset>

**Domain Knowledge:**

E-commerce data is the data that is constantly expanding, creating many possibilities for data analysts to see trends, and understand how the business model is working and what we could do to improve business models.

This dataset, in particular, is unique. It is the largest Pakistan E-commerce data that spans across three years, 2016, 2017 and 2018. It has enough data to provide insights to a business and see how trends have been for the past three years.

**Problem Statement:**

* To understand the trends E-commerce in Pakistan
* To understand customer buying preferences and behavior w.r.t E-commerce in Pakistan

**Data Transformation**

I did data cleaning through PowerBI, but I also did a bit of Data Wrangling through Python as a sample to show the steps I took while cleaning the data before creating my charts.

Some steps include:

* Removing null columns
* Removing null rows (since the data was too large and removing rows did not affect it too much)
* Changing data type of columns such as quantity ordered, month and year.
* Renaming Columns
* Replacing values in status to unify them. For example, the word refund and numerous synonyms that needed to be fixed.

**Dimensions and KPIs:**

“A dimension is a structure that categorizes facts and measures in order to enable users to answer business questions. KPI stands for key performance indicator, a quantifiable measure of performance over time for a specific aim.”

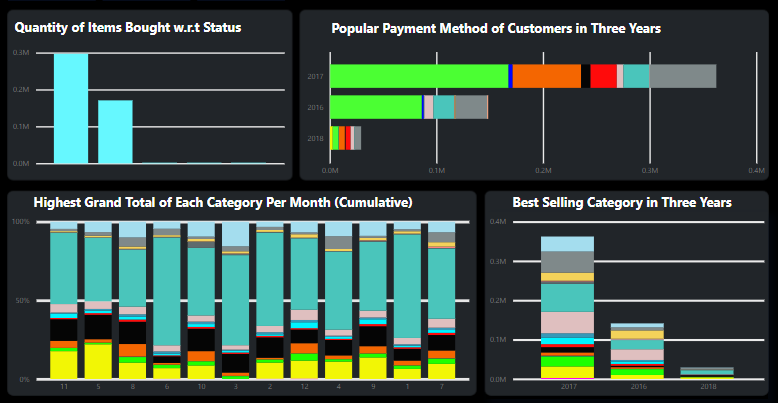
The Dimensions that I have chosen are:

* Time
* Customers
* Payment Method
* Category
* Product
* Status

The KPIs that I have measured are:

* Grand Total
* Quantity of Product Ordered
* Price

**Charts and Explanation**

**Quantity of Items Bought w.r.t Status**

Here the KPI that is being measured is the quantity of items that was ordered with respect to status. Since this is a bar chart. It clearly shows what number of items had a completed status and what amount of items were canceled or refunded. This graph would allow us to look into the business model further as in which items were being canceled or refunded, e.t.c

**Payment Method of Customers in Three Years**

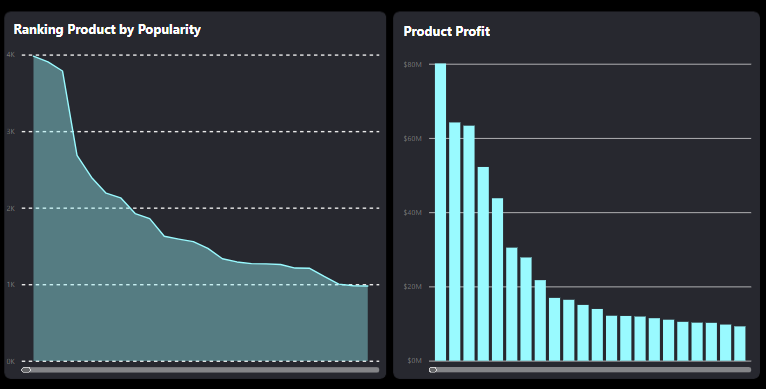
This stacked bar chart is measuring quantity ordered w.r.t year and payment method. This would allow the business to understand which payment method is popular among customers and which one is generating the most revenue so that they would offer discounts and other packages to their customers more accordingly

**Grand Total of Each Category per Month (Cumulative)**

This 100% stacked column chart is measuring the KPI grand total with respect to category and month. This is to analyze which category has being producing the most profit for the business, why is it impacting it in such as way, and why are other categories not doing as well as the most popular one.

**Best Selling Category in Three Years**

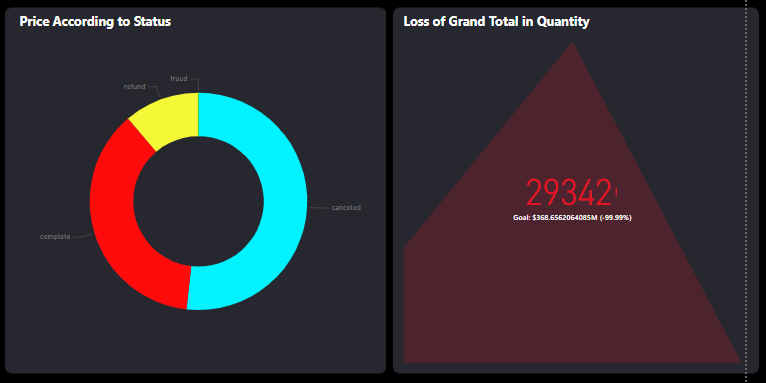
This stacked column chart is measuring the KPI quantity ordered with respect to category and years. This would show which category has been in demand by the customers in the past three years and the business could look into as to why is it so.

**Ranking Product by Popularity**

This area chart measures the KPI quantity ordered by year and product. This chart will show that which item has been ordered the most by customers and how many times. It will give insights on the popularity of each item respectively.

**Product Profit**

This clustered column chart elaborates which product has been giving the most profit in all the three years. Sometimes it is possible that that product may not be the product that is most popular. The chart measures the grand total with respect to grand total to see what revenue is being generated by each item respectively.

**Price According to Status**

This donut chart is measuring the KPI price with respect to status. This would give insight to the business model as to which status has been causing what percentage of loss or gain within the business and then would further explore the reason behind the %ages.

**Loss of Grand Total w.r.t Quantity**

This KPI card illurates the KPIs grand total and quantity throughout the years. It is showing a triangle which means there was in an increase and then a sharp decline through the years which means that the business is not going well in such as case, therefore the business should further investigate why such a thing is occuring.

**Potential Analysis:**

I have written the analysis as queries:

* Which product was the most popular in the last three years?
  + IDROID Jet Black, ordered over 3900 times
* Which product made the most profit in the last three years?
  + MATSAM, giving a grand total of over 80M
* Which category were the customers more inclined to buy in the last three years?
  + Men’s fashion was the most popular in 2016, but was dominated by Mobiles and Tablets' Category in 2017-18. The mobile category had 710000 orders in 2017, which is highest among them all.
* Which method of payment was popular among consumers for the last three years?
  + Most consumers prefer cash on delivery (which is understandable since this data represents the population of Pakistan) however in 2018, by a margin, Payaxis had more popularity.
* How many orders were completed and how many were canceled?
  + Over 230000 were completed and 1700000 canceled
* What category had the highest grand total every month (Cumulative)?
  + Mobiles and Tablets
* What is the % loss caused by cancelation of orders?
  + 51.74% losses are caused by cancelation of orders.

**Other Insights:**

After making the dashboard and creating 6 charts that served the purpose of answering queries, I could notice other patterns within the charts that were interlinked. Some of them were

* In 2017, the people who ordered through Easypay were the people who canceled their orders the most.
* Overall, people with online payment methods tend to cancel their orders more than COD customers
* The mobile and tablet category in 2017 had almost an equal amount of completed and canceled orders despite being the most popular category in 2017.
* The product that gave the highest grand total was actually ordered 2000 times less than the most popular product, but despite that managed to generate a lot of profit.

**Final Thoughts/Opinions**

I personally would prefer python over PowerBI. I felt restricted by the limited options that PowerBI provides relative to Python. Furthermore, the data itself I believe should have had more columns that one could analyze in order to get further insights on the E-commerce data of Pakistan. However, PowerBI had some interesting features. My personal favorite one was the option that highlighted a value in all the charts created to give us hidden/implicit information about the Data, and the variety of charts and visuals that it provides.