AFROPAVO ANALYTICS

CHAKUWEZA COMPANY DATA ANALYSIS REPORT

WILBERT MISINGO

Data Scientist

BACKGROUND

Chakuweza is a business-to-business platform established to enable businesses of all types and sizes to transport goods more efficiently across Tanzania.

The company is focused on building practical solutions for Tanzania's dynamic transportation needs, from developing apps and web solutions, to providing dedicated support for goods on the move.

Chakuweza aggregates a pool of delivery options from 28 ton, 14 ton, 5 ton trucks to pickup trucks, vans and motorcycles.

RESOURCES

Chakuweza Company provided three datasets which were used as a tool for the analysis and drawing a final report, these datasets includes:-

1. **Order delivery details dataset**
   This includes different information concerning the order and its delivery from the point where the order is being set to the point where the order is being received. The dataset consists of 29 columns and 21201 rows of data. The dataset includes data according to the following:-

   - Order No
   - User Id
   - Vehicle Type
   - Platform Type
   - Personal or Business
   - Placement - Day of Month
   - Placement - Weekday (Mo = 1)
   - Placement - Time
   - Confirmation - Day of Month
   - Confirmation - Weekday (Mo = 1)
   - Confirmation - Time
   - Arrival at Pickup - Day of Month
   - Arrival at Pickup - Weekday (Mo = 1)
   - Arrival at Pickup - Time
   - Pickup - Day of Month
   - Pickup - Weekday (Mo = 1)
   - Pickup - Time
   - Arrival at Destination - Day of Month
   - Arrival at Destination - Weekday (Mo = 1)
   - Arrival at Destination - Time
   - Distance (KM)
   - Temperature
   - Precipitation in millimeters
   - Pickup Latitude
   - Pickup Longitude
   - Destination Latitude
   - Destination Longitude
   - Rider Id
   - Time from Pickup to Arrival

2. **Delivery driver/biker details dataset**
   This includes different information about the biker's/driver's identity and his/her performance towards the delivery job. The dataset includes 5 columns and 960 rows. The dataset includes data according to the following:-

   - Rider Id
   - Number Of Orders
   - Age
   - Average Rating
   - Number Of Ratings

3. **Company delivery variables**
   This includes a detailed explanation of each variable used on the previous two datasets and their implication with respect to the company activities.

DATA COLLECTION

During the process, some of the data were missing due to the fact that there seem to be a lost connection on some cases thus further data were required to be gathered to fill in the gap.

From the first dataset (i.e. **Order delivery details dataset**), the following information is provided:-

- Pickup Latitude
- Pickup Longitude
- Destination Latitude
- Destination Longitude

From those information, a python library known as **Reverse Geocode** was used to reverse coding those coordinates into their respective address, thus the following new data were obtained:-

- Pickup Address
- Destination Address

DATA CLEANING AND PROCESSING

Before all the datasets were analyzed, all the datasets were cleaned to produce clean and much more understanding datasets.

During the process, the following columns were removed from the first dataset (i.e. **Order delivery details dataset**), this is due to their lack of connection with the rest of the datasets:-

- "Order No",
- "User Id",
- "Vehicle Type",
- "Placement - Day of Month",
- "Placement - Weekday (Mo = 1)",
- "Placement - Time",
- "Confirmation - Day of Month",
- "Confirmation - Weekday (Mo = 1)",
- "Confirmation - Time",
- "Arrival at Pickup - Day of Month",
- "Arrival at Pickup - Weekday (Mo = 1)",
- "Arrival at Pickup - Time",
- "Pickup - Day of Month",
- "Pickup - Weekday (Mo = 1)",
- "Pickup - Time",
- "Arrival at Destination - Day of Month",
- "Arrival at Destination - Weekday (Mo = 1)",
- "Arrival at Destination - Time",
- "Temperature", "Precipitation in millimeters",
- "Pickup Lat",
- "Pickup Long",
- "Destination Lat",
- "Destination Long"

Thus remaining with the following columns:-

- Pickup Address
- Destination Address
- Platform Type
- Personal or Business
- Distance (KM)
- Rider Id
- Time from Pickup to Arrival

Due to cleaning of the first dataset, a new dataset was created called **“Clean Dataset One”**

Also during the process, the second dataset (i.e. **Delivery driver/biker/rider details dataset**), and the following columns were also removed due to lack of connection with the rest of the data:-

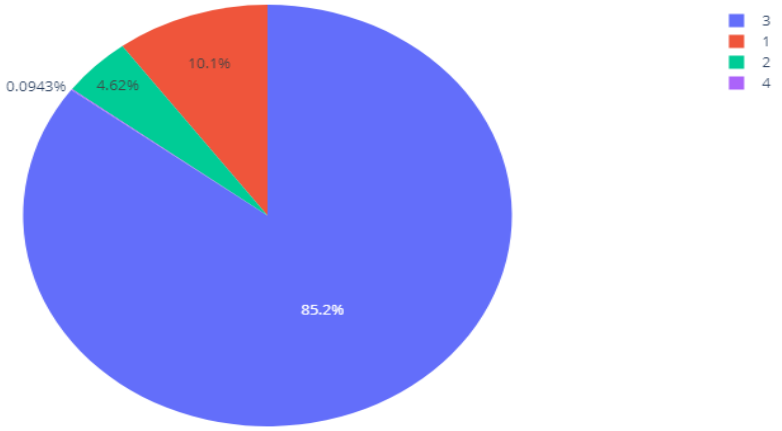
- Average Rating

Thus remaining with the following columns:-

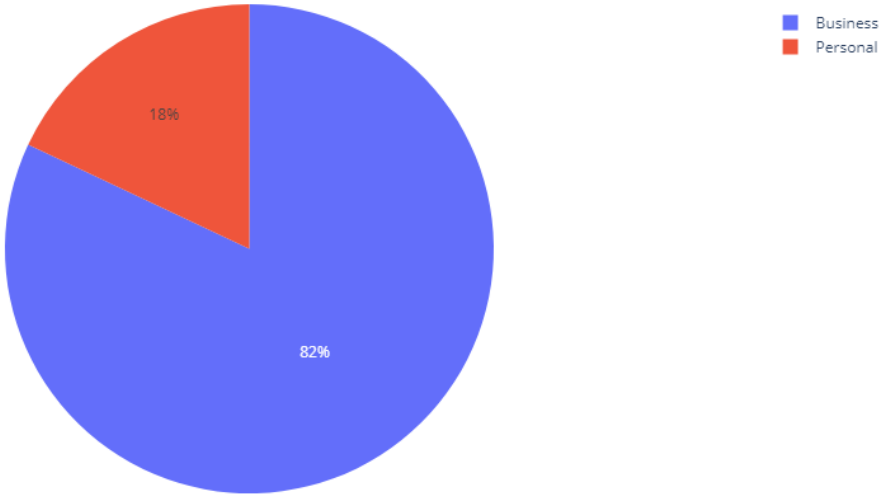
- Rider Id
- Number Of Orders
- Age
- Number Of Ratings

DATA VISUALIZATION

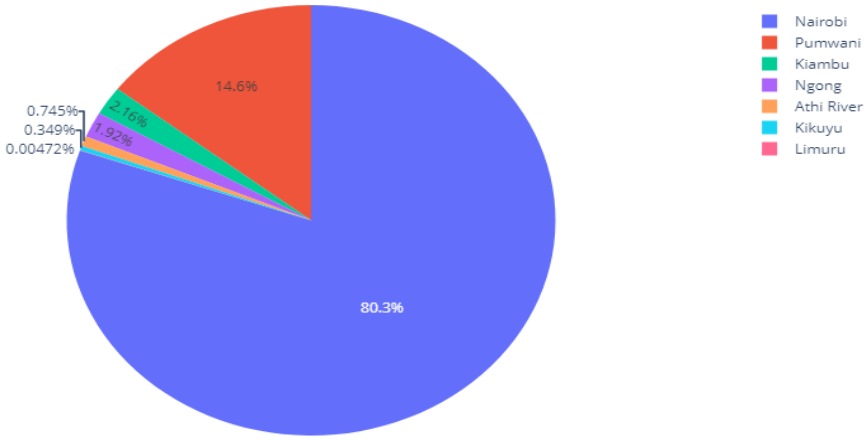
After throughout review of the datasets, different findings were observed according to the following data visualization:-

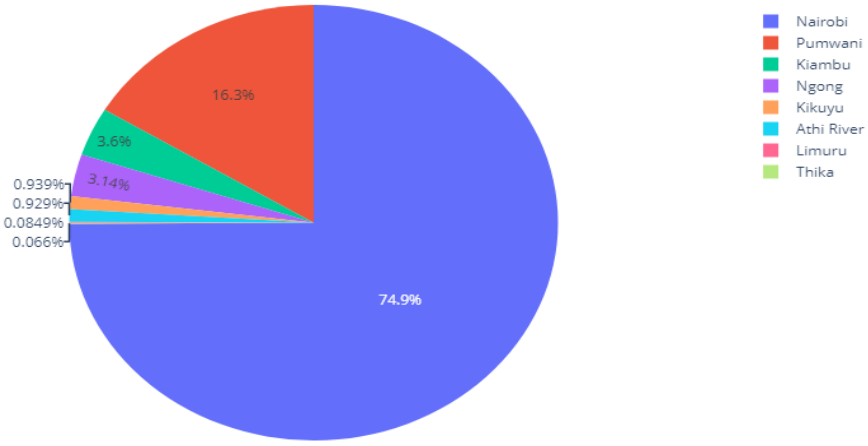


The chart above shows the ratio between different platforms used by the users to make their order for delivery, thus the company should put a lot of effort on Platform 3 since it's the most preferred platform.

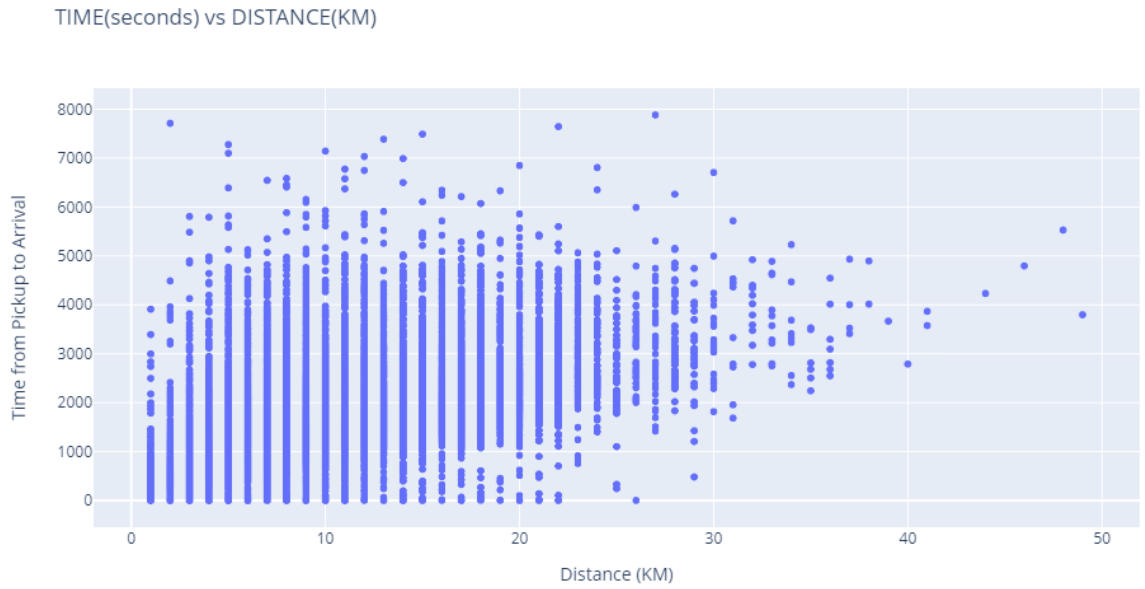


The chart above shows the ratio between the different purposes for making a delivery, most of the delivery products where for business purposes, so the company should invest more on such products by reducing their delivery fee thus to attract more customers.

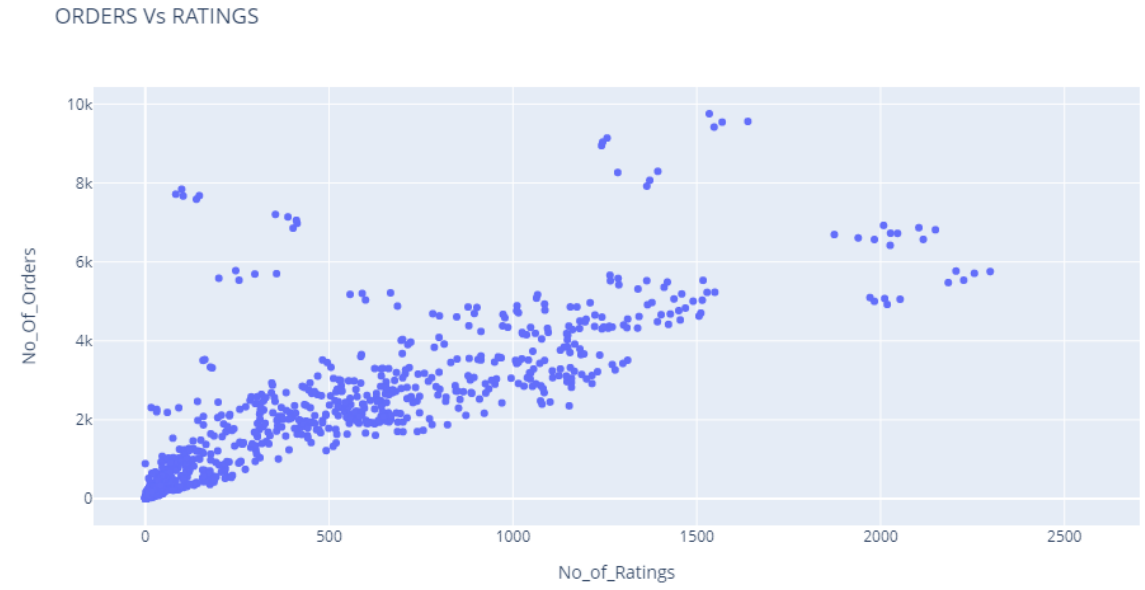




The first chat above is a chart showing the ratio for the addresses for the delivery product pickup while the second shows the ratio for the addresses for the delivery product destination, for all the two charts **Nairobi and Pumwani** have higher rate of deliveries, thus the company should focus more on these regions.



The chart above show the relationship between the distances traveled by a rider to deliver a package in relation to the amount of time it took him/her to deliver that package, there seem to be a variation of time taken by riders to deliver a package on the same distance thus lowering the reputation and production rate of the company.



The chart above shows the relationship between the number of delivery orders made by a rider and the number of rating received from those deliveries, there seem to be a variation due to most of riders making many delivery but receiving little ratings.

RECOMMENDATIONS

Due to the above observations, Chakuweza is recommended to do the following in order to boost their production and trade reputation:-

- Awarding riders who made their delivery faster with higher ratings.
- Suggesting riders with higher ratings for delivery jobs to customers.
- Focusing on “platform 3” as a major platform for delivery orders.
- Focusing on “Nairobi” and “Pumwani” Regions due to their higher delivery traffics.