

KRAFTWER KORP SALES ANALYSIS

Problem Statement

Kraftwerk Korp is Atlas AI's new customer. They have 500 stores spread across Nairobi selling 3 different products: Bits, Bobs and Widgets. Kraftwerk has shared with you as a data analyst their sales data which include the location of each store and total sales volume for each product and expect you as a data analyst to help them answer two key questions:

- 1.What is the relationship between store performance and location?
- 2.Based on that relationship, where should Kraftwerk build more stores?

Data Analysis Guide

We will use the Google Data Analytics framework to guide us through the analysis:

1.Ask - Business Challenge/Objective/Question

The key business questions that Kraftwerk wants to answer include:

- i)What are the top 5 neighborhoods in terms of revenue?
- ii)What are the top 5 stores in terms of sales quantity?
- iii)What is the best selling item in terms of both sales quantity and revenue?
- iv)Which part of Nairobi has the highest concentration of stores?

2.Prepare - Data generation, collection, storage and management

We will load the sales data into R studio session and explore it to understand its structure and properties.

3.Process - Data cleaning and transformation

The data cleaning in this case will involve:

- inspecting for NAs/missing values
- formatting column names
- checking for duplicate records

Data transformation will entail:

- calculating revenue per product
- sorting sales and revenue data
- calculating total revenue and sales for all the products

4.Analyze - Data exploration, visualization and analysis

The analysis phase will involve creating static visualizations such as bar plots, box plots and drawing conclusions.

5.Share - Communicating and interpreting results

We will use tableau tool to help bring the data to life and answer the key business questions.

6. Act - Putting insights to work

Finally, we will make data-driven recommendations to Kraftwerk on store performance and expansion strategies

Part 1: General EDA - Getting to Know the Data

1.1 Importing Required Packages

If the required libraries are not installed in your R studio session, the following script will install and load the libraries

```
#script installs required packages if not installed in the R session
if(!require(rmarkdown)) install.packages("rmarkdown")
if(!require(tinytex)) install.packages("tinytex")
if(!require(tidyverse)) install.packages("tidyverse")
if(!require(skimr)) install.packages("skimr")
if(!require(readr)) install.packages("readr")
if(!require(extrafont)) install.packages("extrafont")
if(!require(extrafont)) install.packages("ggthemes")
if(!require(ggtext)) install.packages("ggtext")

#load the packages
library(rmarkdown)
library(tinytex)
library(tidyverse)
library(skimr)
library(readr)
library(extrafont)
library(ggthemes)
library(ggtext)
#font_import()
#loadfonts(device = "win")
```

1.2 Data Preparation

1.2.1 Load the data

```
data_path = "./www/Kraftwerk_korp_sales.csv"

sales_data = read_csv(data_path)

#script to have the data overview
head(sales_data)
```

```
## # A tibble: 6 x 8
##   id neighborhood shop_name      y      x bits_qty bobs_bqty widgets_qty
##   <dbl> <chr>      <chr>    <dbl> <dbl>   <dbl>   <dbl>      <dbl>
## 1     1  Mlolongo    MG 4 phase 3 d~ 37.0 -1.39    6738    3256      2578
## 2     2  Umoja 2 and 3 Green Grocer S~ 36.9 -1.28    6369    3679       750
## 3     3  Shauri Moyo  RAMANI SHOP 70~ 36.8 -1.30    7681    2052      7361
## 4     4  Buruburu    DAUDI'S ENTERP~ 36.9 -1.28    4473    2961      3466
## 5     5  Kangemi     Seraben superm~ 36.7 -1.27    2629    9458      7218
## 6     6  Kayole      Eagle Services  36.9 -1.28    9706    3896      3865
```

1.2.2 Inspect the data

The sales data has 501 rows, 8 columns and no missing values.

```
#creating function to understand the data structure  
data_structure <- function(df){
```

```
#column data types  
str(df)
```

```
#summary statistics mean,median,max,min  
summary(df)
```

```
#checks for missing values,sd,p0,p25,p50,p75,p100  
skim(df)  
}
```

```
#applying the function to the data  
data_structure(sales_data)
```

```
## spec_tbl_df [501 x 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)  
## $ id          : num [1:501] 1 2 3 4 5 6 7 8 9 10 ...  
## $ neighborhood: chr [1:501] "Mlolongo" "Umoja 2 and 3" "Shauri Moyo" "Buruburu" ...  
## $ shop_name   : chr [1:501] "MG 4 phase 3 dreamland" "Green Grocer Shop" "RAMANI SHOP 7068" "DAUDI'  
## $ y           : num [1:501] 37 36.9 36.8 36.9 36.7 ...  
## $ x           : num [1:501] -1.39 -1.28 -1.3 -1.28 -1.27 ...  
## $ bits_qty    : num [1:501] 6738 6369 7681 4473 2629 ...  
## $ bobs_bqty   : num [1:501] 3256 3679 2052 2961 9458 ...  
## $ widgets_qty : num [1:501] 2578 750 7361 3466 7218 ...  
## - attr(*, "spec")=  
## .. cols(  
## ..   id = col_double(),  
## ..   neighborhood = col_character(),  
## ..   shop_name = col_character(),  
## ..   y = col_double(),  
## ..   x = col_double(),  
## ..   bits_qty = col_double(),  
## ..   bobs_bqty = col_double(),  
## ..   widgets_qty = col_double()  
## .. )  
## - attr(*, "problems")=<externalptr>
```

Table 1: Data summary

Name	df
Number of rows	501
Number of columns	8
Column type frequency:	
character	2
numeric	6
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
neighborhood	0	1	3	29	0	81	0
shop_name	0	1	4	50	0	501	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
id	0	1	251.00	144.77	1.00	126.00	251.00	376.00	501.00	
y	0	1	36.88	0.09	36.63	36.84	36.89	36.92	37.13	
x	0	1	-	0.08	-	-	-	-	-	
			1.26		1.51	1.30	1.27	1.23	1.04	
bits_qty	0	1	4878.81	2938.62	5.00	2302.00	4819.00	7226.00	9995.00	
bobs_bqty	0	1	5175.23	2872.37	4.00	2810.00	5117.00	7710.00	9964.00	
widgets_qty	0	1	4973.45	2905.55	12.00	2351.00	4930.00	7598.00	9994.00	

1.2.1 Checking for duplicates

There are no duplicate data entries in the dataset

```
sum(duplicated(sales_data))
```

```
## [1] 0
```

1.3 Data Cleaning and Transformation

Renaming the columns

```
#renaming columns
sales_data <- sales_data %>%
  rename(
    'shop_location' = neighborhood,
    'lon' = y,
    'lat' = x
  )
head(sales_data)
```

```
## # A tibble: 6 x 8
##   id shop_location shop_name      lon lat bits_qty bobs_bqty widgets_qty
##   <dbl> <chr>      <chr>    <dbl> <dbl>   <dbl>   <dbl>      <dbl>
## 1     1 Mlolongo    MG 4 phase 3 d~ 37.0 -1.39    6738    3256      2578
## 2     2 Umoja 2 and 3 Green Grocer S~ 36.9 -1.28    6369    3679       750
## 3     3 Shauri Moyo  RAMANI SHOP 70~ 36.8 -1.30    7681    2052      7361
## 4     4 Buruburu    DAUDI'S ENTERP~ 36.9 -1.28    4473    2961      3466
## 5     5 Kangemi     Seraben superm~ 36.7 -1.27    2629    9458      7218
## 6     6 Kayole      Eagle Services  36.9 -1.28    9706    3896      3865
```

Calculating the products' revenue

```
sales_revenue_data <- sales_data %>%
  mutate(
    bits_revenue = bits_qty*100,
    bobs_revenue = bobs_bqty*150,
    widgets_revenue = widgets_qty*300,
    total_revenue = bits_revenue+bobs_revenue+widgets_revenue,
```

```

    total_sales = bits_qty+bobs_bqty+widgets_qty
  )
head(sales_revenue_data)

## # A tibble: 6 x 13
##       id shop_location shop_name      lon  lat bits_qty bobs_bqty widgets_qty
##   <dbl> <chr>          <chr>      <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1     1 Mlolongo      MG 4 phase 3 d~ 37.0 -1.39   6738    3256    2578
## 2     2 Umoja 2 and 3 Green Grocer S~ 36.9 -1.28   6369    3679     750
## 3     3 Shauri Moyo  RAMANI SHOP 70~ 36.8 -1.30   7681    2052    7361
## 4     4 Buruburu      DAUDI'S ENTERP~ 36.9 -1.28   4473    2961    3466
## 5     5 Kangemi        Seraben superm~ 36.7 -1.27   2629    9458    7218
## 6     6 Kayole        Eagle Services  36.9 -1.28   9706    3896    3865
## # ... with 5 more variables: bits_revenue <dbl>, bobs_revenue <dbl>,
## #   widgets_revenue <dbl>, total_revenue <dbl>, total_sales <dbl>

```

Saving the processed and cleaned sales data

```
write.csv(sales_revenue_data, "./www/processed_sales_data.csv", row.names = FALSE)
```

calculating cumulative sales by neighborhood

```

sales_revenue_neighborhood <- sales_revenue_data %>%
  group_by(shop_location) %>%
  summarise(
    neighborhood_revenue = sum(total_revenue)
  )
head(sales_revenue_neighborhood)

```

```

## # A tibble: 6 x 2
##   shop_location neighborhood_revenue
##   <chr>                <dbl>
## 1 Athi River          20498750
## 2 Bahati              1077650
## 3 Buruburu            20370850
## 4 CBD                 2779500
## 5 Chokaa              4526700
## 6 Dagoretti Corner    8101350

```

Part 2: General EDA - Answering Business Questions

We will set a plotting theme for our plots

```

#setting theme
f1<-"Comic Sans MS"

custom_theme<-function()
{
  #setting font
  Font<-"Comic Sans MS"

  theme_economist() %+replace%

  theme(
    #####grid elements#####

```

```

panel.grid.major = element_blank(), #strip major gridlines
panel.grid.minor = element_blank(), #strip minor gridlines
axis.ticks = element_blank(), #strip axis ticks
axis.line = element_blank(),
#####text elements#####
plot.title = element_markdown( #title
  family = Font, #set font family
  size = 13, #set font size
  face = 'bold', #bold typeface
  hjust = 0, #left align
  vjust = 2), #raise slightly

plot.subtitle = element_markdown( #subtitle
  family = Font, #font family
  size = 12,hjust = 0), #font size

plot.caption = element_text( #caption
  family = Font, #font family
  size = 8.7,
  hjust = 0
), #right align

axis.title = element_text( family = Font,size = 9), #font size

axis.text = element_text( family = Font,size = 9), #font size

legend.text = element_text( family = Font,size = 9),
legend.title = element_text( family = Font,size = 9),
legend.position = "right",
plot.background = element_rect(fill = "white", color = NA)

#####plot#####
)
}

```

Q1 - What are the top 5 neighborhoods in terms of revenue?

```

#filtering top 5 neighborhoods in terms of revenue
top5_neighborhoods <- sales_revenue_neighborhood %>%
  arrange(desc(neighborhood_revenue)) %>%
  top_n(5)

```

```
## Selecting by neighborhood_revenue
```

```
head(top5_neighborhoods)
```

```

## # A tibble: 5 x 2
##   shop_location neighborhood_revenue
##   <chr>             <dbl>
## 1 Thika             78133900
## 2 Kayole            76688100
## 3 Kibera            50975000

```

```
## 4 Huruma          39947950
## 5 Ruiru            39075400
```

Plotting top 5 neighborhoods in terms of revenue

```
top5_neighborhoods %>%
  ggplot(aes(x=reorder(shop_location,neighborhood_revenue),y=neighborhood_revenue))+
  geom_col(fill="#ffc107")+
  labs(title = "Top 5 neighborhoods in terms of revenue",y="",x="Neighborhood")+
  geom_text(aes(label=scales::comma(neighborhood_revenue,prefix = "Ksh")),color="black",
            hjust=0.6,vjust=0.5,family=f1,size=2.8)+
  coord_flip()+
  custom_theme()+
  theme(
    axis.text.x = element_blank()
  )
```

