VIRGINIA COMMONWEALTH UNIVERSITY

STATISTICAL ANALYSIS & MODELING

A5: VISUALIZATION – PERCEPTUAL MAPPING FOR BUSINESS USING PYTHON AND R

Ahad Zifain Miyanji
V01108270

Date of Submission: 17/07/2024

CONTENTS

VISUALIZATION – PERCEPTUAL MAPPING FOR BUSINESS USING PYTHON AND R

INTRODUCTION

This report analyzes the dataset "NSSO68.csv" using statistical models to derive meaningful insights and relationships. The methodologies explored include logistic regression, decision tree analysis, and Tobit regression. Logistic regression is used to predict binary outcomes by modeling the probability of a categorical response variable based on one or more predictor variables. Decision trees offer a non-linear approach by recursively partitioning data into subsets, making them interpretable and suitable for complex decision-making processes. Tobit regression addresses scenarios involving censored data, where the dependent variable is constrained by upper or lower limits, providing robust estimation methods essential in fields like economics and healthcare. By comparing these methodologies, this report aims to demonstrate their applicability and effectiveness in extracting valuable insights from the "NSSO68.csv" dataset, thereby contributing to informed decision-making in various research and practical domains

.OBJECTIVES

- Plot a histogram to show the distribution of total consumption across different districts and a bar plot to visualize consumption per district with district names for the Chhattisgarh state.

- Plot any variable on the Chhattisgarh state map using NSSO68.csv data

BUSINESS SIGNIFICANCE

Visualization and perceptual mapping are powerful tools for businesses to gain insights and make informed decisions. Visualization simplifies complex data into easy-to-understand graphical formats, helping identify trends, patterns, and outliers. It enhances communication with stakeholders and

supports data-driven decision-making while enabling real-time performance monitoring. Perceptual mapping, on the other hand, allows businesses to understand their market positioning relative to competitors by visualizing consumer perceptions. It helps identify market gaps, informs competitive analysis, guides product development, and shapes effective marketing strategies. Together, these tools provide a comprehensive understanding of market dynamics and consumer preferences, leading to better strategic planning and execution

RESULTS AND INTERPRETATION

a) Plot a **histogram** (to show the distribution of total consumption across different districts) and a **barplot** (To visualize consumption per district with district names) to indicate the consumption district-wise for the Chattisgarh state.

R Code:

```
# Summarize consumption
chtsdnew$total_consumption <- rowSums(chtsdnew[, c("ricepds_v", "Wheatpds_q", "chicken_q",
"pulsep_q", "wheatos_q")], na.rm = TRUE)

# Summarize and display top consuming districts and regions
summarize_consumption <- function(group_col) {
summary <- chtsdnew %>%
group_by(across(all_of(group_col))) %>%
summarise(total = sum(total_consumption)) %>%
arrange(desc(total))   return(summary)
}

district_summary <- summarize_consumption("District") region_summary
<- summarize_consumption("Region")

cat("Top Consuming Districts:\n")
print(head(district_summary, 4)) cat("Region
Consumption Summary:\n")
print(region_summary)

# Rename districts and sectors district_mapping <- c("1" = "Koriya","2" = "Surguja","3" =
"Jashpur","4" = "Raigarh","5" = "Korba","6" = "Janjgir - Champa","7" = "Bilaspur","8" =
"Kawardha","9" = "Rajnandgaon","10" =
"Durg","11" = "Raipur","12" = "Mahasamund","13" = "Dhamtari","14" = "Kanker","15" =
"Bastar","16" = "Dantewada","17" = "Narayanpur","18" = "Bijapur") sector_mapping
<- c("2" = "URBAN", "1" = "RURAL")

chtsdnew$District <- as.character(chtsdnew$District) chtsdnew$Sector
<- as.character(chtsdnew$Sector)
chtsdnew$District <- ifelse(chtsdnew$District %in% names(district_mapping),
district_mapping[chtsdnew$District], chtsdnew$District)
```

```
chtsdnew$Sector <- ifelse(chtsdnew$Sector %in% names(sector_mapping),
sector_mapping[chtsdnew$Sector], chtsdnew$Sector)

View(chtsdnew)

hist(chtsdnew$total_consumption, breaks = 10, col = 'blue', border = 'black',
     xlab = "Consumption", ylab = "Frequency", main = "Consumption Distribution in Chattisgarh
State")

CHTSD_consumption <- aggregate(total_consumption ~ District, data = chtsdnew, sum)
View(CHTSD_consumption)
??barplot
barplot(CHTSD_consumption$total_consumption,
names.arg = CHTSD_consumption$District,
las = 2, # Makes the district names vertical
col = 'blue',          border = 'black',          xlab =
"District",          ylab = "Total Consumption",
main = "Total Consumption per District",
        cex.names = 0.6) # Adjust the size of district names if needed
```

Python Code:

```
 # Summarize consumption
chtsdnew['total_consumption'] = chtsdnew[['ricepds_v', 'Wheatpds_q', 'chicken_q', 'pulsep_q', 'wheatos
_q']].sum(axis=1)

# Summarize and display top consuming districts and regions def summarize_consumption(df,
group_col):     summary =
df.groupby(group_col)['total_consumption'].sum().reset_index().sort_values(by='total_co nsumption',
ascending=False)     return summary

district_summary = summarize_consumption(chtsdnew, 'District') region_summary
= summarize_consumption(chtsdnew, 'Region')

print("Top Consuming Districts:\n")
print(district_summary.head(4)) print("Region
Consumption Summary:\n")
print(region_summary)
```

Results

```
Top Consuming Districts:

    District  total_consumption
10        11        1530.338289
9         10        1503.413102
1          2        1367.023922
6          7         967.668297
Region Consumption Summary:

   Region  total_consumption
1       2        8800.393114
2       3        2369.273470
0       1        1817.004637
```

```
Top Consuming Districts:
> print(head(district_summary, 4))
# A tibble: 4 × 2
  District total
     <int> <dbl>
1       11 1530.
2       10 1503.
3        2 1367.
4        7  968.
> cat("Region Consumption Summary:\n")
Region Consumption Summary:
> print(region_summary)
# A tibble: 3 × 2
  Region total
   <int> <dbl>
1      2 8800.
2      3 2369.
3      1 1817.
```

Interpretation:

The above figures show the summary of the districts and region based on total consumption. The top consuming districts are displayed here. District with the most consumption is 11 with 1530 units. The same is with region. With region 2 being the highest of 8800 units.

b) Plot for any variable on the **Chattisgarh** state map using NSSO68.csv data

R Code:

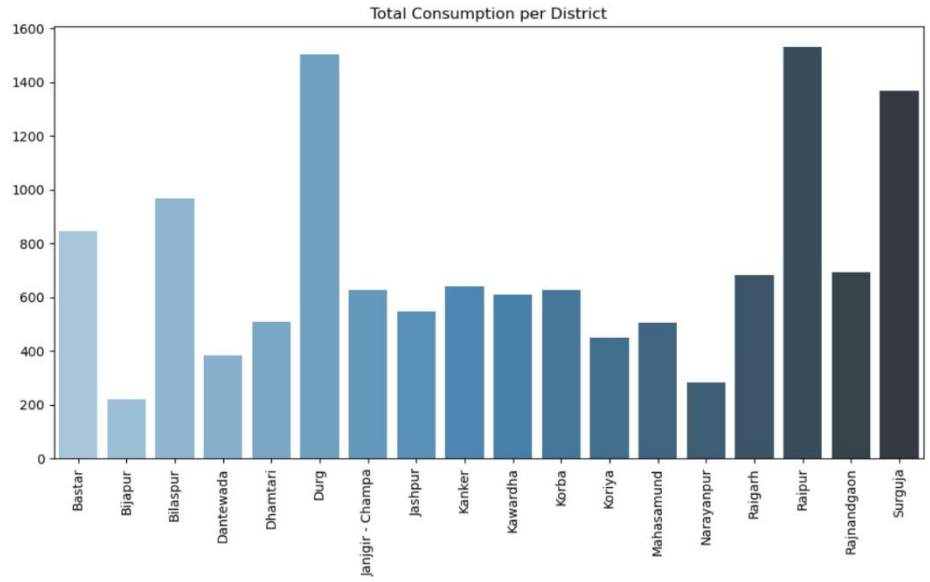
```
data_map <- st_read("E:\\R\\Assignment 5\\CHHATTISGARH_DISTRICTS.geojson")  View(data_map)


data_map <- data_map %>%
rename(District = dtname)
colnames(data_map)
data_map_data <- merge(CHTSD_consumption,data_map,by = "District")
View(data_map_data) ggplot(data_map_data) +    geom_sf(aes(fill
=total_consumption, geometry = geometry)) +    scale_fill_gradient(low =
"yellow", high = "red") +    ggtitle("Total Consumption_by_District")


ggplot(data_map_data) +    geom_sf(aes(fill = total_consumption, geometry =
geometry)) +    scale_fill_gradient(low = "yellow", high = "red") +    ggtitle("Total
Consumption by District") +   geom_sf_text(aes(label = District, geometry =
geometry), size = 3, color = "black")
```

Python Code:

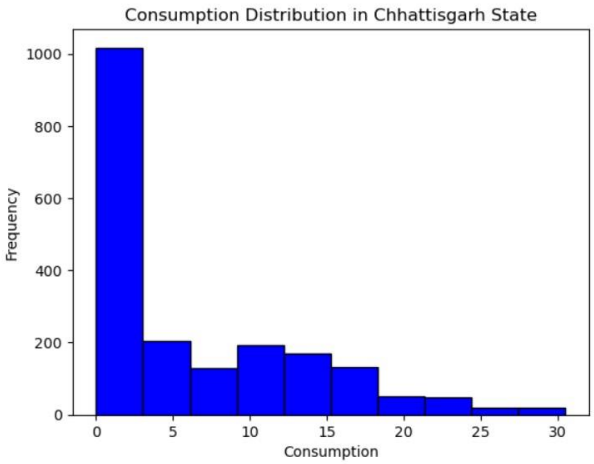
```
# Plot using GeoPandas and Matplotlib fig, ax = plt.subplots(1, 1, figsize=(15, 10))
data_map_data.plot(column='total_consumption', ax=ax, legend=True,
cmap='YlOrRd') plt.title("Total Consumption by District") plt.show()
```

```
# Annotate districts with their names fig,
ax = plt.subplots(1, 1, figsize=(15, 10))
data_map_data.plot(column='total_consu
mption', ax=ax, legend=True,
cmap='YlOrRd')
data_map_data.apply(lambda x:
ax.annotate(text=x['District'],
xy=x.geometry.centroid.coords[0],
ha='center'), axis=1) plt.title("Total
Consumption by District") plt.show()
```

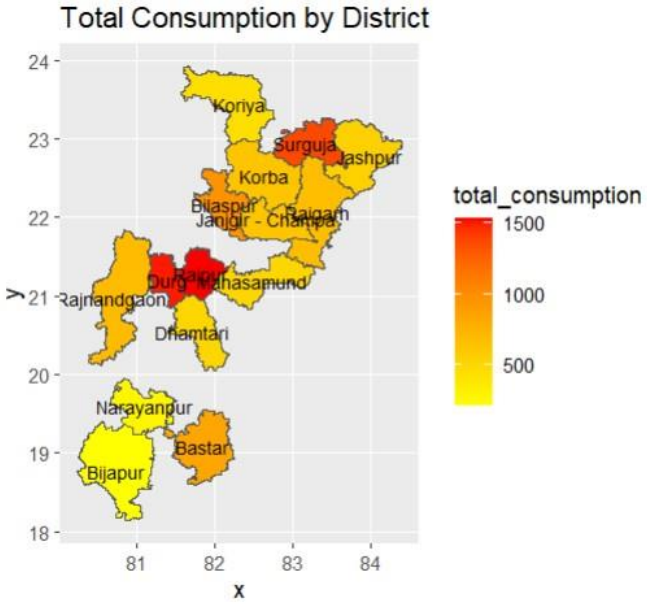
Result:



Bar Plot



Histogram



Interpretation:

Histogram of Consumption Distribution in Chhattisgarh State

The histogram provides a visual representation of the distribution of consumption levels across Chhattisgarh State. The x-axis indicates the varying levels of consumption, while the y-axis shows the frequency of each consumption level. From the histogram, it is apparent that the majority of the population falls within the lower consumption levels, with a noticeable decrease in frequency as consumption levels rise. This indicates that most individuals in Chhattisgarh consume relatively little, while a small fraction of the population exhibits significantly higher consumption levels. The data suggest a skewed distribution, where the bulk of consumption values are clustered at the lower end, and very few individuals are at the higher end of the spectrum. This pattern highlights economic inequality within the state, pointing to a considerable gap between the majority with low consumption and the minority with high consumption.

Bar Chart of Total Consumption per District

The bar chart presents a comparison of total consumption levels across various districts in Chhattisgarh. Each bar in the chart corresponds to a specific district, with the height of the bar representing the total consumption for that district. The chart demonstrates substantial variability in consumption levels among the districts. Notably, Durg district emerges with the highest total consumption, closely followed by Raipur and Surguja.

Conversely, districts such as Bijapur and Narayanpur register significantly lower total consumption levels. This disparity underscores the economic inequalities and varying consumption patterns across the state, suggesting that districts like Durg, Raipur, and Surguja may have greater resource availability or more favorable economic conditions compared to districts like Bijapur and Narayanpur. The bar chart effectively highlights these differences, providing insights into regional economic statuses.

Choropleth Map of Total Consumption by District

The choropleth map offers a geographical depiction of total consumption per district within Chhattisgarh. The map uses color-coding to represent consumption levels, with a gradient from yellow indicating lower consumption to red indicating higher consumption. This visual tool vividly showcases the disparities in consumption levels across the state. Districts such as Durg, Raipur, and Surguja, shaded in red, are identified as having higher consumption levels. In contrast, districts like Bijapur and Narayanpur are shaded in yellow, highlighting their lower consumption levels. The choropleth map not only provides a spatial understanding of consumption distribution but also reveals regional inequalities. This geographic representation is crucial for informing policy decisions and resource allocation strategies, enabling targeted interventions in lower consumption districts to bridge the economic divide.

Conclusion

These visualizations collectively offer a thorough analysis of consumption patterns within Chhattisgarh. The histogram reveals a skewed consumption distribution, with the majority of the population at the lower end. The bar chart and choropleth map emphasize significant regional disparities, showing that some districts, such as Durg, Raipur, and Surguja, exhibit much higher consumption levels compared to others like Bijapur and Narayanpur. These insights are invaluable for government and organizational decision-making, guiding strategies to address resource distribution, foster economic development, and implement targeted interventions in districts with lower consumption levels. By understanding these patterns, efforts can be better directed towards achieving a more equitable economic landscape across Chhattisgarh.