VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A5.b : Visualisation – Perceptual Mapping for Business

SAYA SANTHOSH

V01101901

Date of Submission: 15-07-2024

CONTENTS

Introduction - Plotting the variable ‘nontotal_v’ on Karnataka state map

This study aims to visually represent the patterns of meat consumption in the state of Karnataka by utilizing data obtained from the National Sample Survey Office (NSSO). Our objective is to graph the variable 'nontotal_v,' which represents the total intake of meat, over the map of Karnataka state using the NSSO68 dataset. This visualization aims to depict the spatial distribution and magnitude of meat consumption across various districts in Karnataka. The NSSO68 dataset offers extensive data on diverse consumption-related aspects for both rural and urban sectors. Through the mapping of the 'nontotal_v' variable, our objective is to discern and emphasize regions characterized by both elevated and diminished levels of meat consumption. This analysis will offer useful insights to policymakers and stakeholders, allowing them to gain a better understanding of regional consumption trends in Karnataka. This will enable them to make well-informed decisions about targeted interventions and resource allocation.

Objectives :

- Visualize meat consumption trends.
- Identify regional consumption patterns.

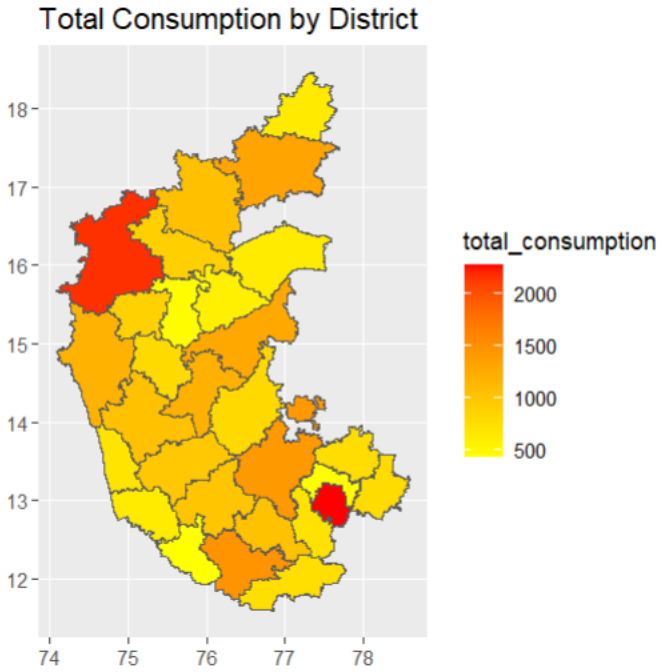
Business Significance :

This holds significant business importance. It offers vital information on the consumption trends in different regions, which helps firms make strategic decisions and allocate resources effectively. By mapping meat consumption, firms may pinpoint regions with significant demand and optimize supply chains to ensure streamlined distribution and save logistical expenses. This data also facilitates market segmentation endeavors, allowing businesses to

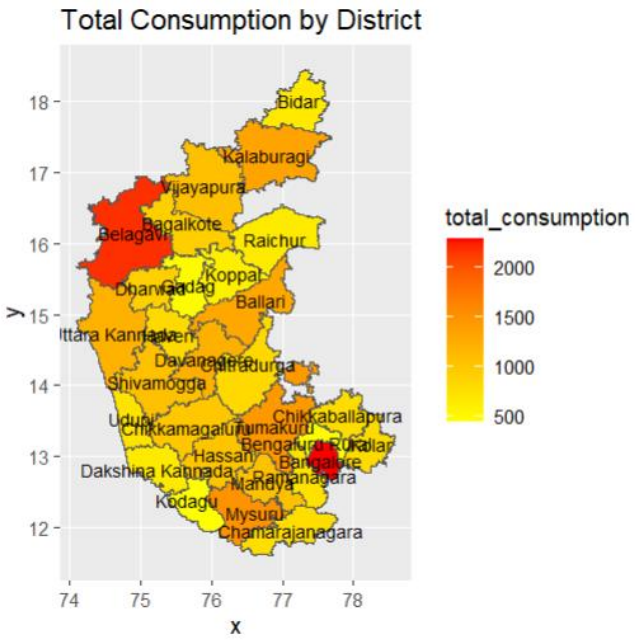
customize marketing methods and product offerings according to district-specific preferences. Furthermore, the map assists policymakers in evaluating dietary patterns and health consequences, directing efforts to encourage well-rounded nutrition and public health consciousness. The geo map assignment promotes well-informed corporate plans, improves market competitiveness, and aids in sustainable development and consumer welfare in Karnataka. Multidimensional Scaling (MDS).

Results and Interpretation using R

```
> # Plot without labeling district names
> ggplot(data_map_data) +
+   geom_sf(aes(fill =total_consumption, geometry = geometry)) +
+   scale_fill_gradient(low = "yellow", high = "red") +
+   ggtitle("Total Consumption by District")
```




```
> # Plot with labelled district names
> 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, col
or = "black")
```

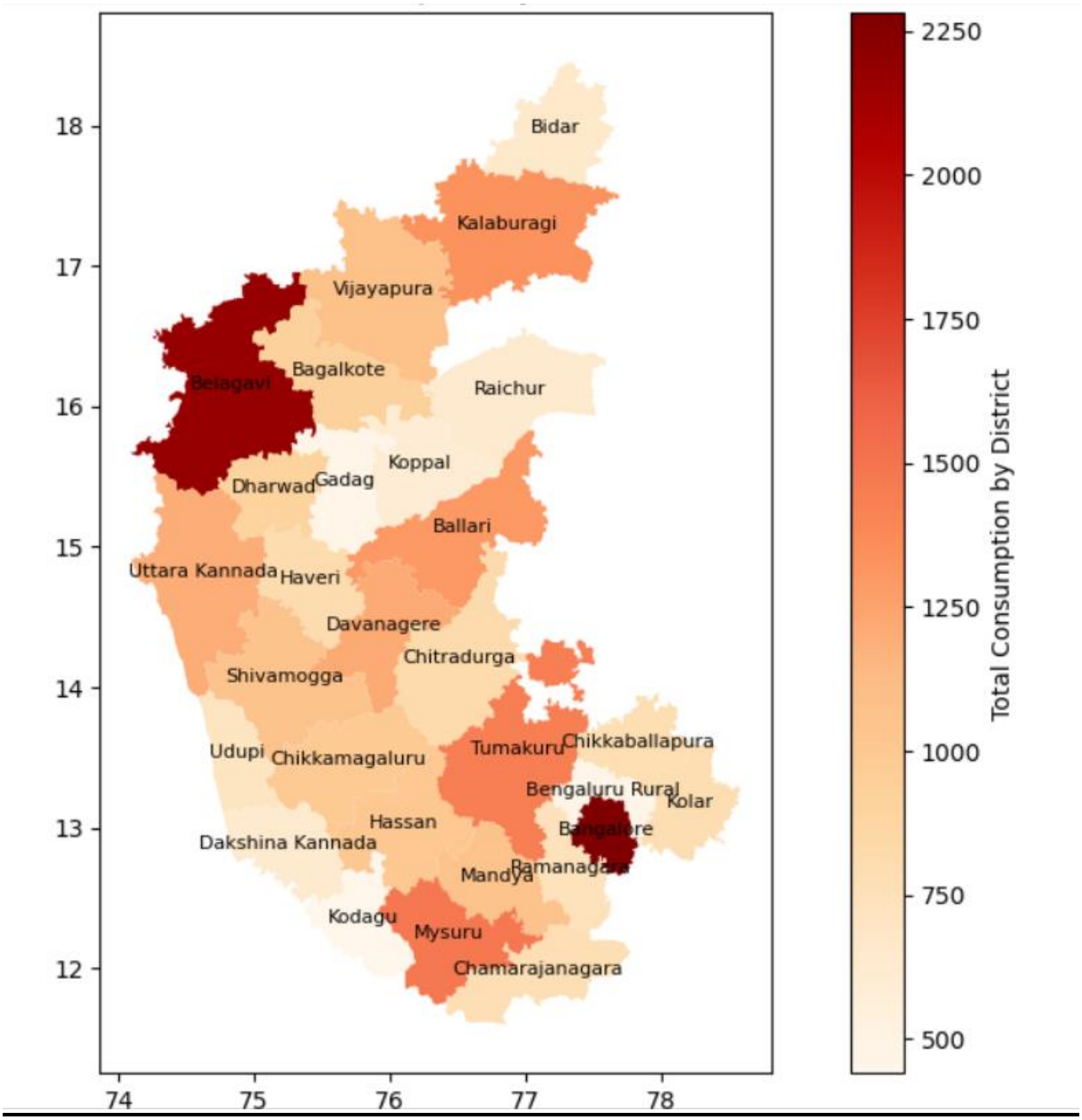


Interpretation:

The map shows total consumption by district in the state, with colors ranging from yellow to red indicating increasing levels of consumption. Districts such as Belagavi and Bengaluru Urban, shaded in red, have the highest total consumption, exceeding 2000 units. In contrast, districts like Kodagu and Chamarajanagara, shaded in yellow, have lower consumption levels around 500 units. This indicates significant regional disparities in consumption patterns, with urban and more densely populated areas exhibiting higher consumption compared to rural areas.

Results and Interpretation using Python

```
# Plot with labeled district names
fig, ax = plt.subplots(1, 1, figsize=(12, 8))
data_map_data.plot(column='total_consumption', cmap='OrRd', legend=True, ax=ax, legend_kwds={'label': "Total Consumption by District"})
data_map_data.apply(lambda x: ax.annotate(text=x['District'], xy=x.geometry.centroid.coords[0], ha='center', fontsize=8, color='black'), axis=1)
plt.title('Total Consumption by District')
plt.show()
```



Interpretation:

The map depicts the total consumption by district in Karnataka, India, with darker shades indicating higher consumption levels. Belagavi and Bengaluru are the districts with the highest total consumption, as indicated by the darkest red shades. Other districts such as Mysuru and Kalaburagi also show relatively high consumption, but to a lesser extent. In contrast, districts like Bidar, Koppal, and Chamarajanagara have the lowest total consumption, as shown by the lightest shades on the map. This gradient provides a visual representation of the disparity in consumption levels across different regions in Karnataka.

Recommendations

Based on the consumption patterns shown in the map of Karnataka, several recommendations can be made to promote balanced and sustainable development. In high-consumption districts like Belagavi and Bengaluru, focus on improving infrastructure and resource allocation, while introducing sustainability initiatives to manage resources effectively. For lower consumption districts such as Bidar, Koppal, and Chamarajanagara, invest in economic development initiatives and support for agriculture and rural entrepreneurship. Enhance infrastructure in rapidly growing areas, encourage economic diversification, and launch public awareness campaigns on sustainable practices. Establish a monitoring and evaluation system to track consumption patterns and policy effectiveness, ensuring data-driven decision-making across the state.

R Codes

```
# b) Plotting total consumption on the Karnataka state map


# Filtering for Karnataka
df_ka <- data %>%
  filter(state_1 == "KA")

# Sub-setting the data
ka_new <- df_ka %>%
  select(state_1, District, Region, Sector, State_Region, Meals_At_Home, ricepds_v,
Wheatpds_q, chicken_q, pulsep_q, wheatos_q, No_of_Meals_per_day)

# Check for missing values in the subset
```

```
cat("Missing Values in Subset:\n")
print(colSums(is.na(ka_new)))

# Impute missing values with mean for specific columns
ka_new$Meals_At_Home <- impute_with_mean(ka_new$Meals_At_Home)

# Check for missing values after imputation
cat("Missing Values After Imputation:\n")
print(colSums(is.na(ka_new)))

# Finding outliers and removing them
outlier_columns <- c("ricepds_v", "chicken_q")
for (col in outlier_columns) {
  ka_new <- remove_outliers(ka_new, col)
}

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

district_summary <- summarize_consumption("District")
cat("District Consumption Summary:\n")
print(district_summary)

# mapping districts so that meging of the tables will be easier
district_mapping <- c(
  "1" = "Belagavi",
  "2" = "Bagalkote",
  "3" = "Vijayapura",
  "4" = "Kalaburagi",
  "5" = "Bidar",
  "6" = "Raichur",
  "7" = "Koppal",
  "8" = "Gadag",
```

```
  "9" = "Dharwad",
  "10" = "Uttara Kannada",
  "11" = "Haveri",
  "12" = "Ballari",
  "13" = "Chitradurga",
  "14" = "Davanagere",
  "15" = "Shivamogga",
  "16" = "Udupi",
  "17" = "Chikkamagaluru",
  "18" = "Tumakuru",
  "19" = "Kolar",
  "20" = "Bangalore",
  "21" = "Bengaluru Rural",
  "22" = "Mandya",
  "23" = "Hassan",
  "24" = "Dakshina Kannada",
  "25" = "Kodagu",
  "26" = "Mysuru",
  "27" = "Chamarajanagara",
  "28" = "Ramanagara",
  "29" = "Chikkaballapura"
)

ka_new$District <- as.character(ka_new$District)
ka_new$District <- district_mapping[ka_new$District]
#ka_new$District <- ifelse(ka_new$District %in% names(district_mapping), district_
mapping[ka_new$District], ka_new$District)
View(ka_new)


# ka_consumption stores aggregate of total consumption district wise
ka_consumption <- aggregate(total_consumption ~ District, data = ka_new, sum)

View(ka_consumption)
```

```
#Plotting total consumption on the Karnataka state

Sys.setenv("SHkaE_RESTORE_SHX" = "YES")

data_map <- st_read("C:\\Users\\sayas\\Downloads\\KARNATAKA_DISTRICTS.geojson")
View(data_map)

data_map <- data_map %>%
  rename(District = dtname)

# merging ka_consumption and data_map tables
data_map_data <- merge(ka_consumption,data_map,by = "District")
View(data_map_data)

# Plot without labeling district names
ggplot(data_map_data) +
  geom_sf(aes(fill =total_consumption, geometry = geometry)) +
  scale_fill_gradient(low = "yellow", high = "red") +
  ggtitle("Total Consumption by District")

# Plot with labelled district names
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 Codes

```
# b) Plotting total consumption on the Karnataka state map
```

```
# Filtering for Karnataka
df_ka = data[data['state_1'] == "KA"]

# Sub-setting the data
ka_new = df_ka[['state_1', 'District', 'Region', 'Sector', 'State_Region', 'Meals_At_Ho
me', 'ricepds_v', 'Wheatpds_q', 'chicken_q', 'pulsep_q', 'wheatos_q', 'No_of_Meals_per
_day']]
# Check for missing values in the subset
print("Missing Values in Subset:")
print(ka_new.isnull().sum())

# Impute missing values with mean for specific columns
ka_new['Meals_At_Home'].fillna(ka_new['Meals_At_Home'].mean(), inplace=True)

# Check for missing values after imputation
print("Missing Values After Imputation:")
print(ka_new.isnull().sum())
# Remove outliers
for col in outlier_columns:
    ka_new = remove_outliers(ka_new, col)
# Summarize consumption
ka_new['total_consumption'] = ka_new[['ricepds_v', 'Wheatpds_q', 'chicken_q', 'pulse
p_q', 'wheatos_q']].sum(axis=1)

district_summary = ka_new.groupby('District')['total_consumption'].sum().reset_inde
x().sort_values(by='total_consumption', ascending=False)
print("District Consumption Summary:")
print(district_summary)
# Mapping districts so that merging of the tables will be easier
district_mapping = {
    "1": "Belagavi", "2": "Bagalkote", "3": "Vijayapura", "4": "Kalaburagi", "5": "Bida
r",
    "6": "Raichur", "7": "Koppal", "8": "Gadag", "9": "Dharwad", "10": "Uttara Kanna
da",
```

```
    "11": "Haveri", "12": "Ballari", "13": "Chitradurga", "14": "Davanagere", "15": "Sh
ivamogga",
    "16": "Udupi", "17": "Chikkamagaluru", "18": "Tumakuru", "19": "Kolar", "20": "
Bangalore",
    "21": "Bengaluru Rural", "22": "Mandya", "23": "Hassan", "24": "Dakshina Kanna
da",
    "25": "Kodagu", "26": "Mysuru", "27": "Chamarajanagara", "28": "Ramanagara", "
29": "Chikkaballapura"
}

ka_new['District'] = ka_new['District'].astype(str).map(district_mapping).fillna(ka_ne
w['District'])
print(ka_new)
# ka_consumption stores aggregate of total consumption district-wise
ka_consumption = ka_new.groupby('District')['total_consumption'].sum().reset_index
()
print(ka_consumption)
# Load and plot Karnataka state map
data_map = gpd.read_file("C:\\Users\\sayas\\Downloads\\KARNATAKA_DISTRICT
S.geojson")

data_map = data_map.rename(columns={'dtname': 'District'})
print(data_map)
# Merging ka_consumption and data_map tables
data_map_data = data_map.merge(ka_consumption, on='District')
print(data_map_data)
# Plot with labeled district names
fig, ax = plt.subplots(1, 1, figsize=(12, 8))
data_map_data.plot(column='total_consumption', cmap='OrRd', legend=True, ax=ax,
legend_kwds={'label': "Total Consumption by District"})
data_map_data.apply(lambda x: ax.annotate(text=x['District'], xy=x.geometry.centroi
d.coords[0], ha='center', fontsize=8, color='black'), axis=1)
plt.title('Total Consumption by District')
plt.show()
```

References

1. www.github.com