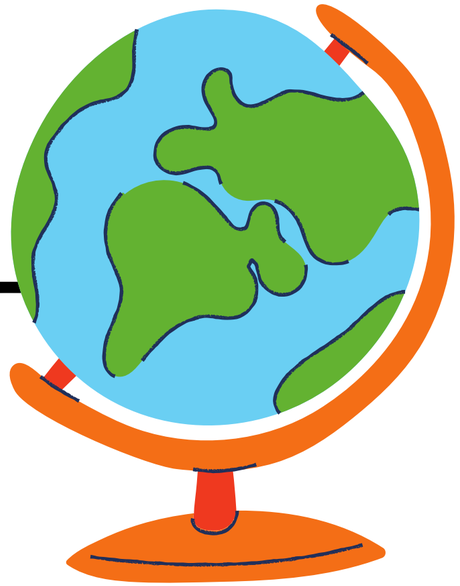
BUILDING THE TOURIST PURCHASING POWER INDEX (TPPI)

Unveiling the true cost of being a tourist worldwide

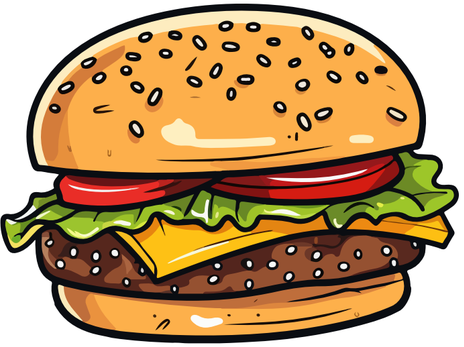


By Meeth Jaswani
3036183937

PROBLEM DEFINITION

Using a Single ITEM?

- Use Coke Price? Use Big Mac price? (Big Mac Index)
- Else use a basket of Factors? -**Rent, Internet, Taxi, Food**
- Creating the Tourist Purchasing Power Index(TPPI)

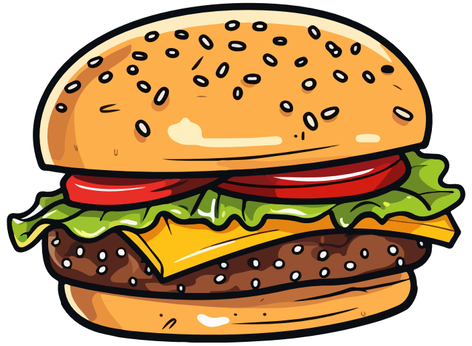


PROBLEM DEFINITION

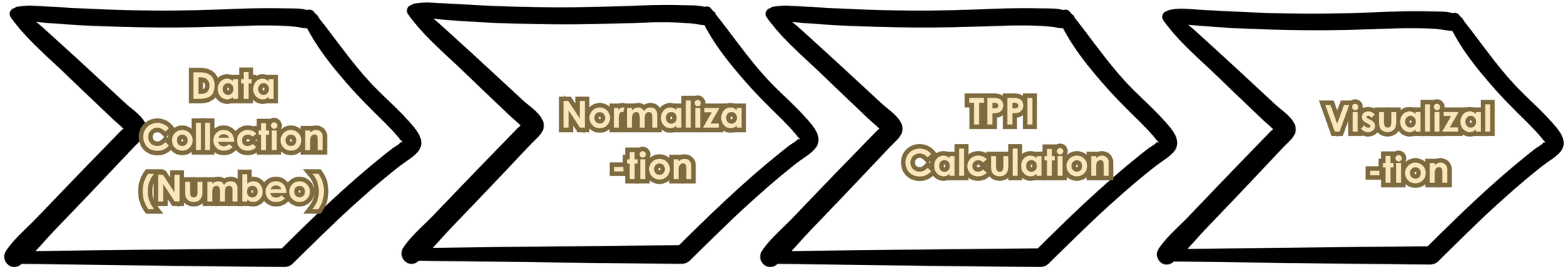
Why Data Science?

- Intuition can **mislead** tourists — what "feels cheap" may not reflect actual affordability
- This approach helps quantify affordability across countries using **objective** metrics

- **Expected Outcome:**
- How real-life tourist expenses vary across countries
- Coke Prices -very rough proxy for tourist affordability



LET'S BUILD THE INDEX!



DATA COLLECTION & CLEANING

NUMBEO 's Data involving the following factors:

1. Rent for a 1-bedroom apartment in the city center
2. Monthly internet cost (60 Mbps or more)
3. Taxi fare per kilometer
4. Cost of a meal at an inexpensive restaurant
5. Price of Coke/Pepsi (0.33 liter bottle)

Data was collected, cleaned(Remove commas and dollar signs) and made into a Single Data frame

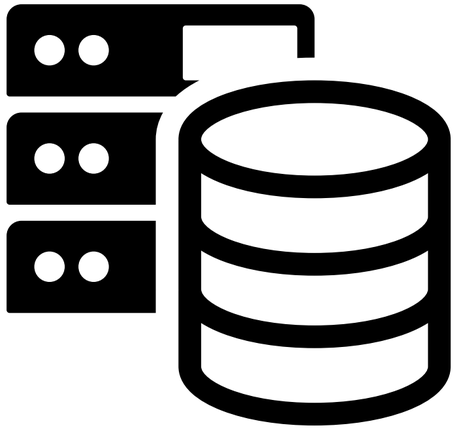
NORMALIZING DATA

Why Normalize?

- Categories like rent and internet are on different scales
- Without adjustment, expensive items (e.g. rent) would skew results

Min-Max Normalization:

$$\text{Normalized Value} = \frac{\text{Actual Price} - \text{Min Price}}{\text{Max Price} - \text{Min Price}}$$



TPPI CALCULATION

Index Formula:

$$\text{Tourist PPP Index} = \frac{\text{Normalized Rent} + \text{Internet} + \text{Taxi} + \text{Meal}}{4}$$

Final Interpretation:

$$\text{Tourist Purchasing Power Index} = 1 - \text{Tourist PPP Index}$$

TPPP Index: Tourist Purchasing Power Parity Index

TPPI : Tourist Purchasing Power Index

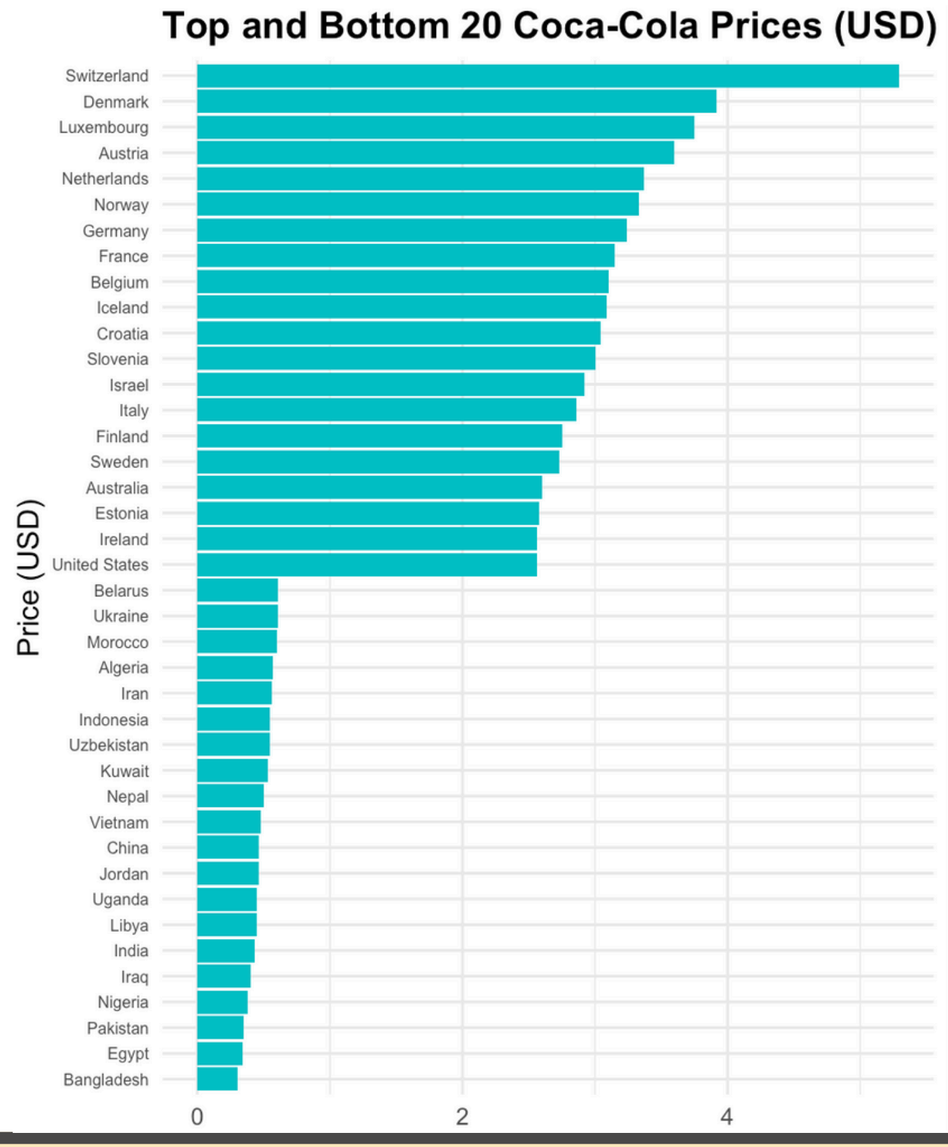
The higher the TPPI, the more bang you get for your tourist buck.

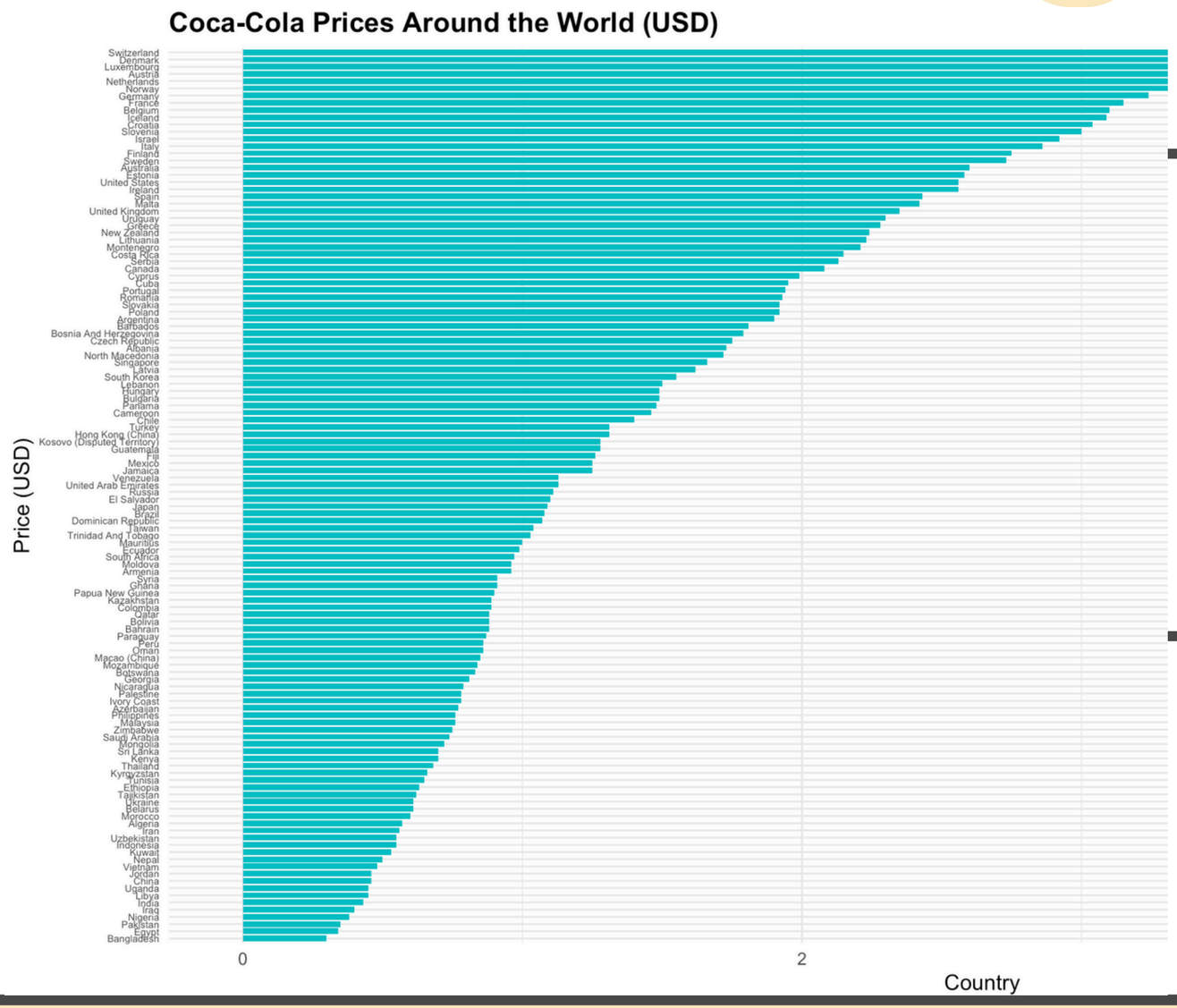
```r
# --- Combine all 4 categories ---
combined_df <- bind_rows(meal_df, rent_df, internet_df, taxi_df)

# Normalize prices per category
normalized_df <- combined_df |>
  group_by(Category) |>
  mutate(Normalized = (Price - min(Price)) / (max(Price) - min(Price))) |>
  ungroup()

# Pivot to wide format
wide_df <- normalized_df |>
  select(Country, Category, Normalized) |>
  pivot_wider(names_from = Category, values_from = Normalized)

# Compute indexes
wide_df <- wide_df |>
  mutate(
    Tourist_PPP = rowMeans(select(., Meal, Rent, Internet, Taxi), na.rm = TRUE),
    Tourist_PPI = 1 - Tourist_PPP
  )
```

VISUALIZATIONS AND CODE

THE COCA-COLA INDEX

A Quick Glance at Global Affordability

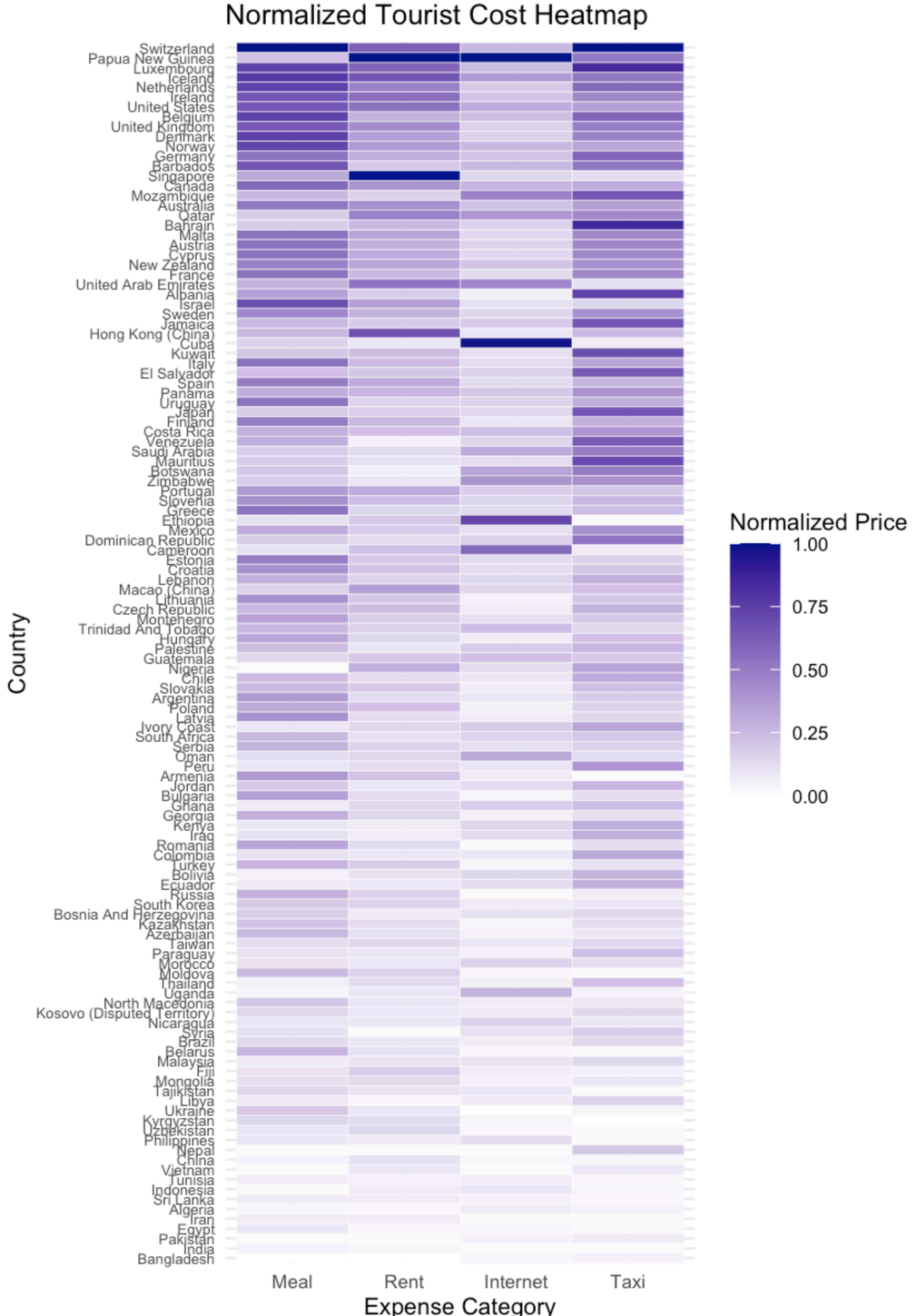




```r
# --- Visual 1: Coca-Cola Prices (Dot Plot) ---
ggplot(coke_df, aes(x = reorder(Country, Price), y = Price)) +
  geom_bar(stat = "identity", fill = "#00BFC4", width=0.8) +
  coord_flip() +  # Flip the coordinates to make bars horizontal
  theme_minimal() +  # Use minimal theme
  theme(
    axis.text.y = element_text(size = 5),  # Adjust y-axis text size
    plot.title = element_text(size = 14, face = "bold")  # Title styling
  ) +
  labs(
    title = "Coca-Cola Prices Around the World (USD)",
    x = "Price (USD)",
    y = "Country"
  )
```

```r
# Get top 20 and bottom 20 countries based on Coca-Cola price
top_bottom_coke_df <- coke_df %>%
  arrange(desc(Price)) %>%  # Sort by Price in descending order for top 20
  slice_head(n = 20) %>%  # Select top 20 countries
  bind_rows(
    coke_df %>%
      arrange(Price) %>%  # Sort by Price in ascending order for bottom 20
      slice_head(n = 20)  # Select bottom 20 countries
  )

# Plot the top 20 and bottom 20 countries by Coca-Cola price
ggplot(top_bottom_coke_df, aes(x = reorder(Country, Price), y = Price)) +
  geom_bar(stat = "identity", fill = "#00BFC4") +  # Adjust bar width
  coord_flip() +  # Flip coordinates for horizontal bars
  theme_minimal() +  # Minimal theme
  theme(
    axis.text.y = element_text(size = 6),  # Adjust y-axis text size
    plot.title = element_text(size = 14, face = "bold")  # Title styling
  ) +
  labs(
    title = "Top and Bottom 20 Coca-Cola Prices (USD)",
    x = "Price (USD)",
    y = "Country"
  )
```

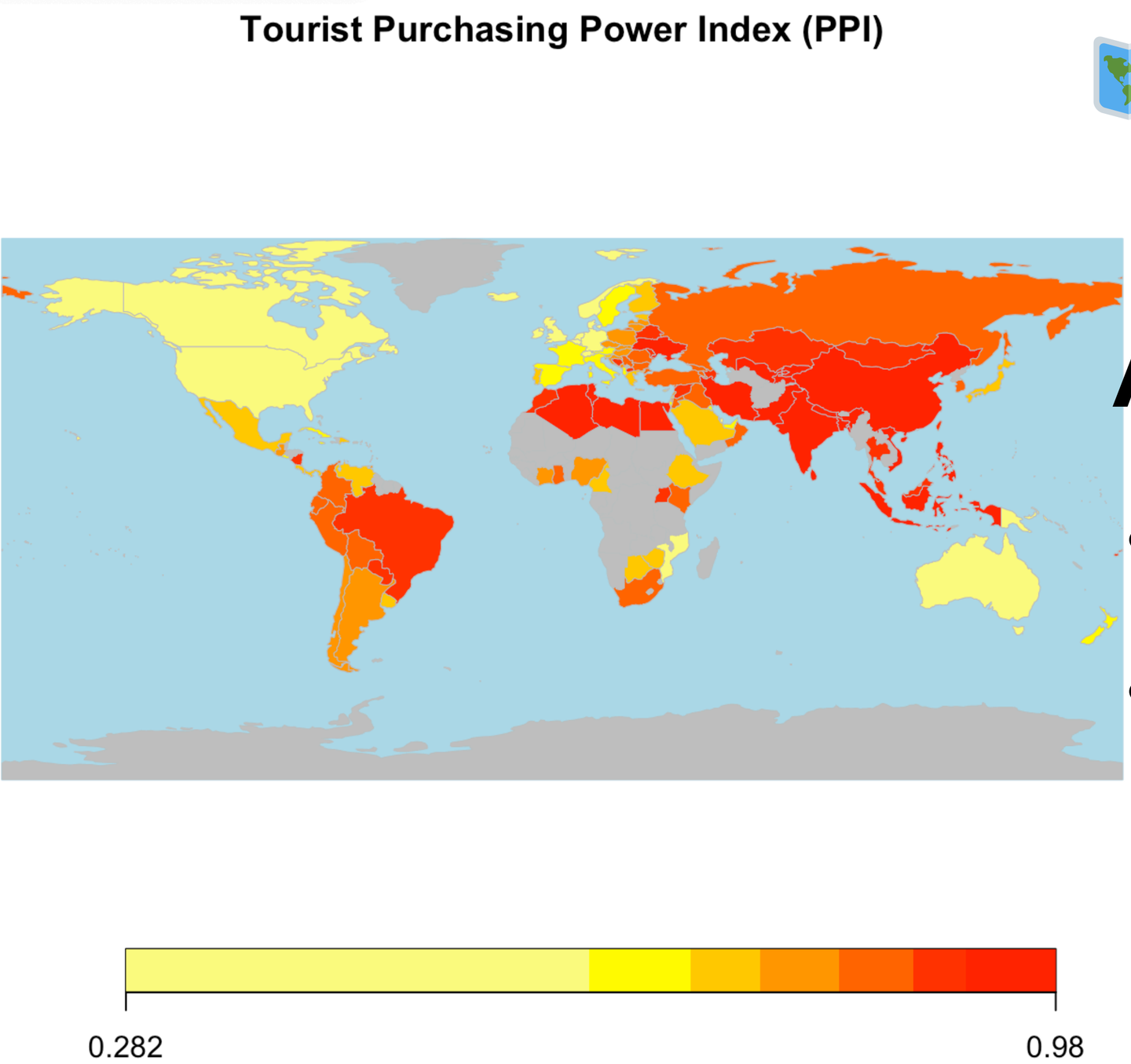
🌡️ TPP INDEX: HEATMAP VIEW



A higher TPPI score means more tourist-friendly pricing.

```r
# --- Visual 2: Normalized Heatmap ---
melted <- wide_df |>
  select(Country, Tourist_PPI, Meal, Rent, Internet, Taxi) |>
  melt(id.vars = c("Country", "Tourist_PPI"))

ggplot(melted, aes(x = variable, y = reorder(Country, -Tourist_PPI), fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "white", high = "darkblue") +
  theme_minimal(base_size = 8) +
  labs(title = "Normalized Tourist Cost Heatmap",
       x = "Expense Category",
       y = "Country",
       fill = "Normalized Price") +
  theme(axis.text.y = element_text(size = 5))  # Adjust y-axis label size here
```

WORLD VIEW OF TOURIST AFFORDABILITY

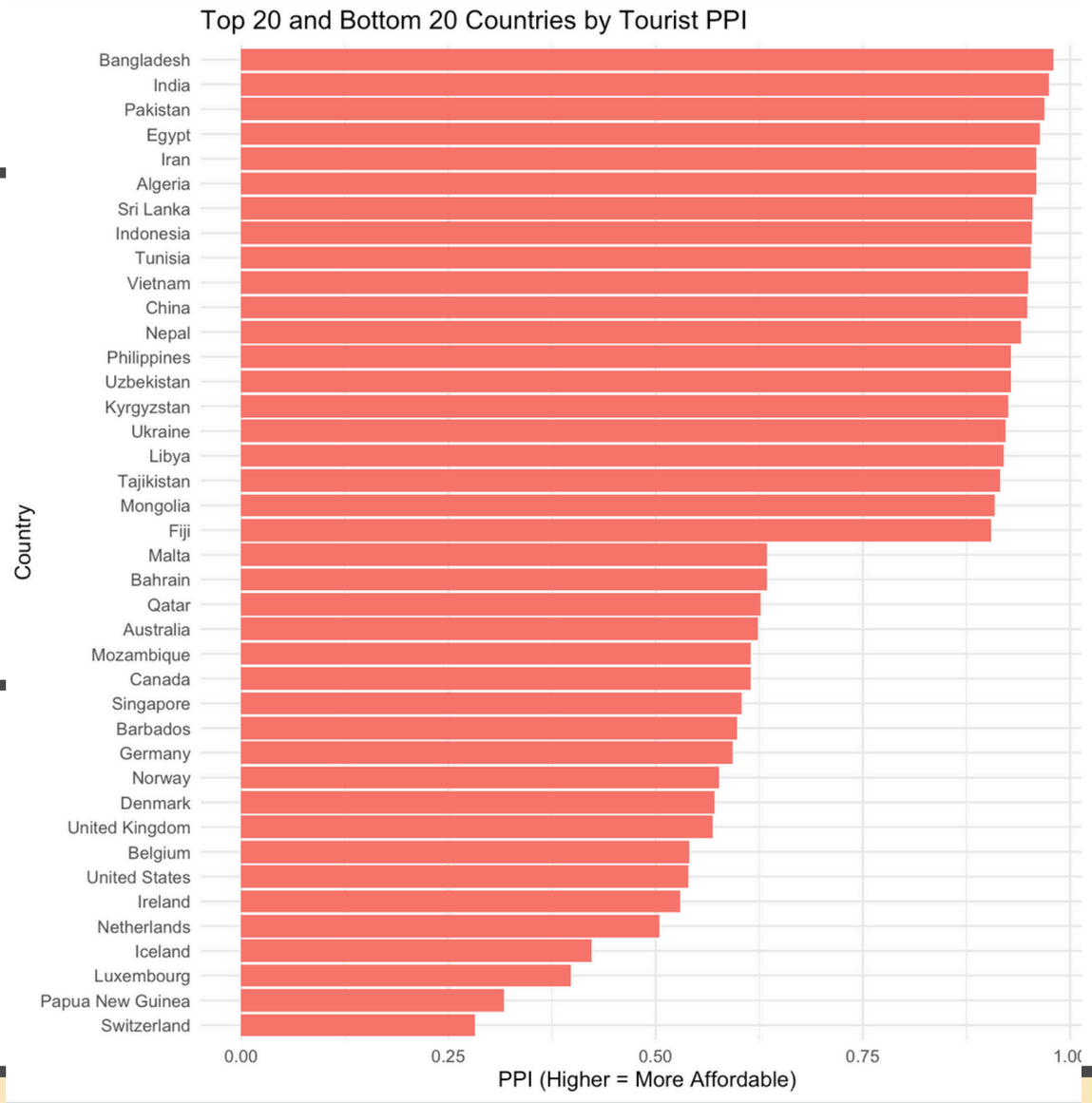


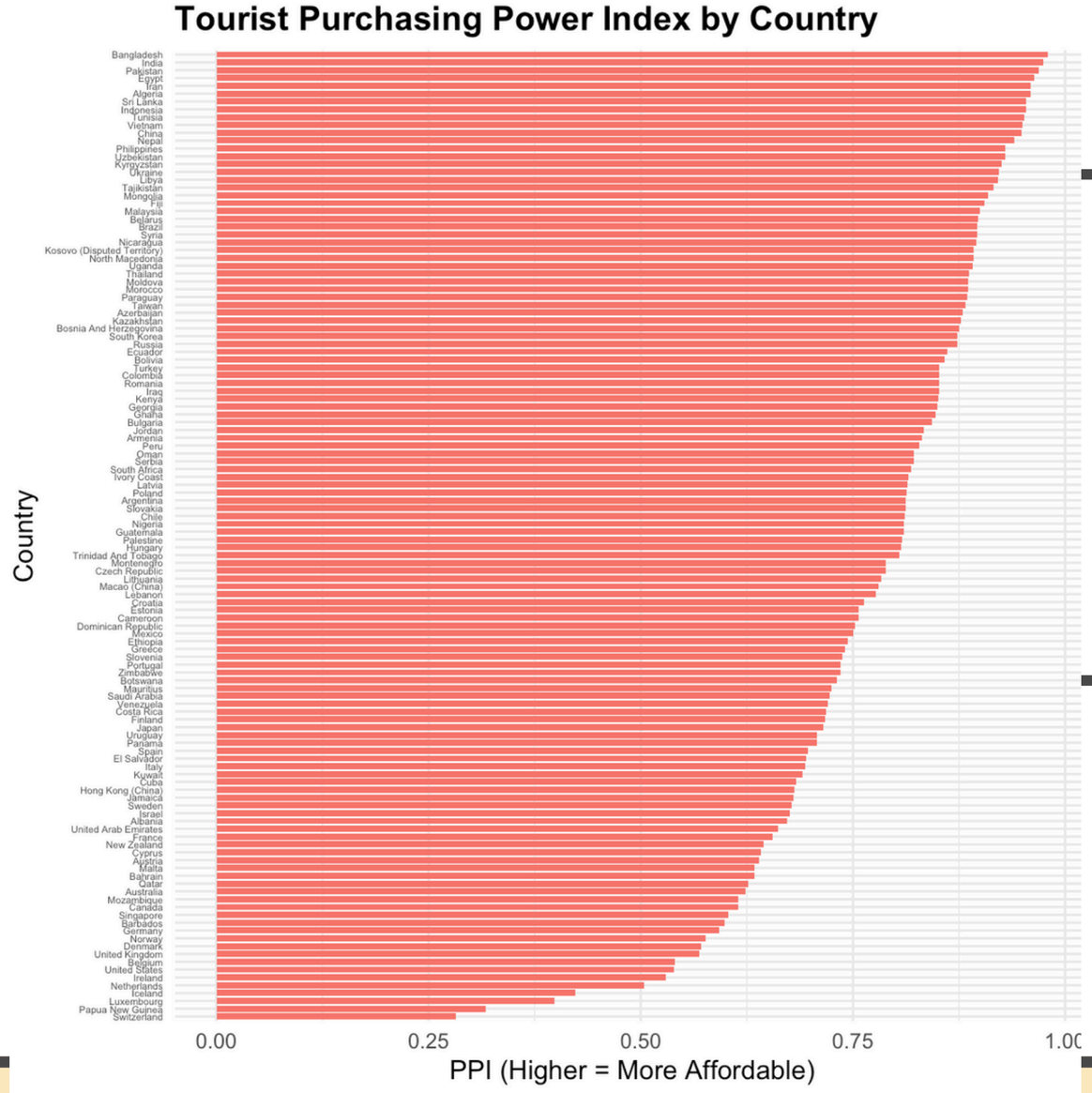
- Darker shades = More affordable for tourists.
- Libraries used-> rworldmap,countrycode

```r
# --- Visual 3: World Map of Tourist PPI ---
map_data <- wide_df |>
  mutate(iso3 = countrycode(Country, "country.name", "iso3c")) |>
  select(iso3, Tourist_PPI)

map_df <- joinCountryData2Map(map_data, joinCode = "ISO3", nameJoinColumn = "iso3")
mapCountryData(map_df, nameColumnToPlot = "Tourist_PPI",
               mapTitle = "Tourist Purchasing Power Index (PPI)",
               colourPalette = "heat", oceanCol = "lightblue", missingCountryCol = "grey")
```

THE TPPI INDEX

A glance at Tourist affordability



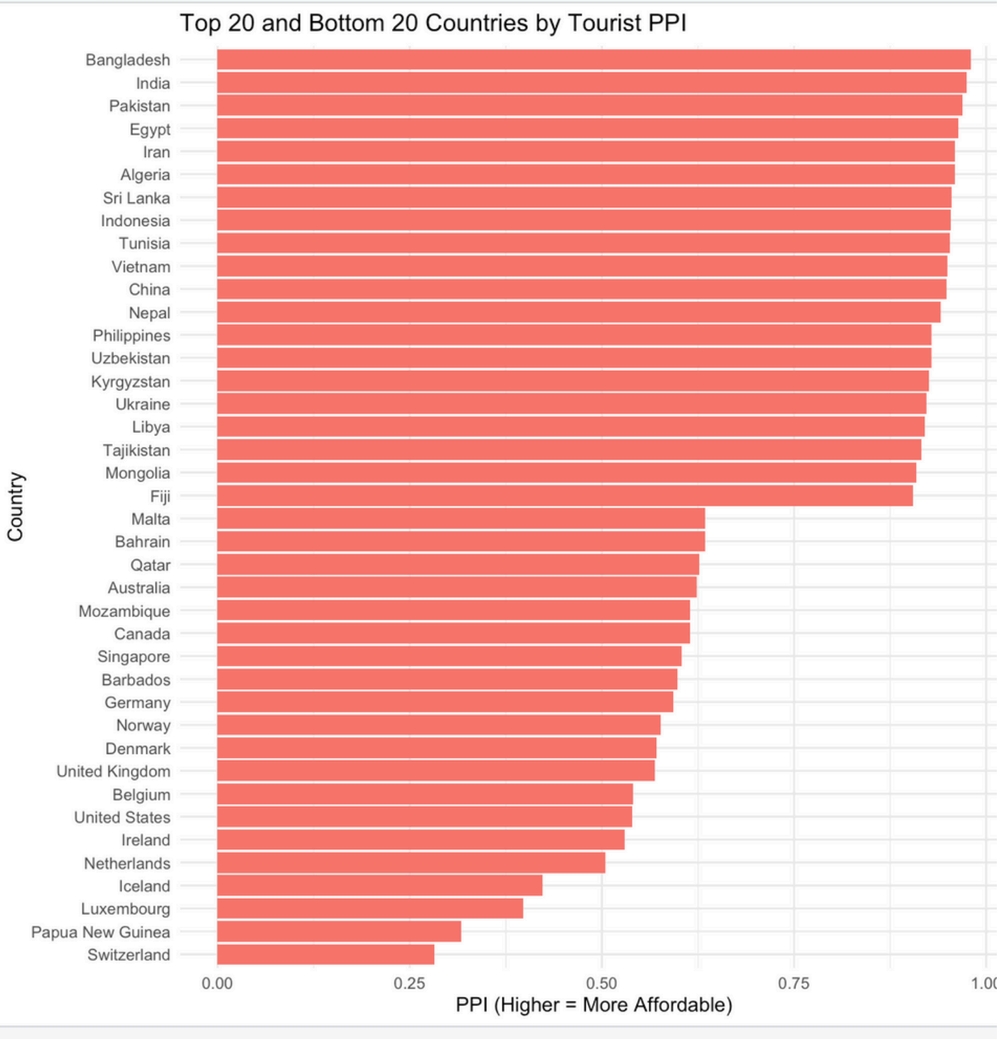


```r
# --- Visual 4: PPI Barplot for All Countries ---
ggplot(wide_df, aes(x = reorder(Country, Tourist_PPI), y = Tourist_PPI)) +
  geom_bar(stat = "identity", fill = "#F8766D", width=0.8) +  # Use geom_bar with stat="identity"
  coord_flip() +  # Flip the coordinates to make bars horizontal
  theme_minimal() +  # Use minimal theme
  theme(
    axis.text.y = element_text(size = 4),  # Adjust y-axis text size
    plot.title = element_text(size = 14, face = "bold")  # Title styling
  ) +
  labs(title = "Tourist Purchasing Power Index by Country", x = "Country", y = "PPI (Higher = More Affordable)")
```

```r
# Get top 20 and bottom 20 countries based on Tourist PPI
top_bottom_df <- wide_df |>
  arrange(Tourist_PPI) |>
  slice_head(n = 20) |>  # Select bottom 20 countries
  bind_rows(
    wide_df |>
      arrange(desc(Tourist_PPI)) |>
      slice_head(n = 20)  # Select top 20 countries
  )

# Plot using geom_bar
ggplot(top_bottom_df, aes(x = reorder(Country, Tourist_PPI), y = Tourist_PPI)) +
  geom_bar(stat = "identity", fill = "#F8766D") +
  coord_flip() +
  theme_minimal(base_size = 9) +
  labs(title = "Top 20 and Bottom 20 Countries by Tourist PPI", x = "Country", y = "PPI (Higher = More Affordable)")
```

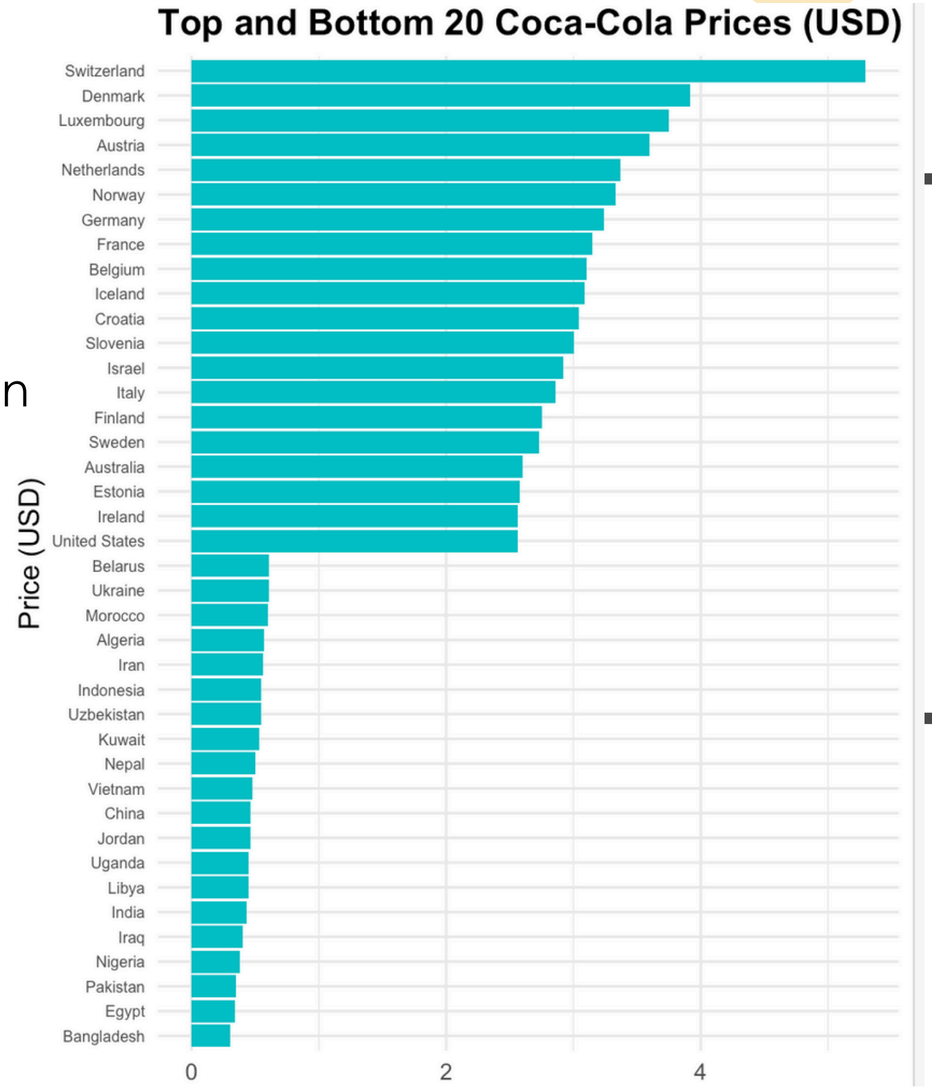
CONCLUSION

THE UNEXPECTED TRUTH BEHIND COCA-COLA & TPPI



Key Observations:

- Bangladesh Cheapest in Both
- Swistherland Most expensisive in both
- Several Countries overlapping



WHY DOES THIS MATTER?

- Coca-Cola Index: A fun first look, but lacks depth.
- TPPI: The real deal for tourists, giving a more complete picture.
- Takeaway: The Coca-Cola price can give us a hint, but the TPPI truly reveals the cost of living for tourists.

THE TWIST: FROM FUN METRIC TO POWERFUL TOOL

What's Next?

- Expanding the TPPI to include more factors (healthcare, entertainment, etc.)
- Not relying on Numbeo as the sole dataset as its skewed towards western nations
- Developing an interactive tool for tourists to customize affordability based on their travel preferences.

THANK YOU