

# Mapping Wage Disparities in the Philippines

## Target Audience

This project is designed for policymakers, researchers, and the general public seeking insights into wage disparities across Philippine regions, areas, industries, education, and gender. The specific needs of each are as follows:

- **Policymakers** need summarized information to guide labor policies and economic planning.
- **Researchers** need detailed and easy to access visualizations to conduct analyses of labor market trends across different characteristics.
- **The general public** needs intuitive representations to easily understand wage trends.

These users all require clear and accessible insights into wage disparities to understand labor trends and inform decisions.

In providing these needs, we potentially face the following challenges:

- **Balancing detail and simplicity.** The visualization must be detailed but not overwhelming to users. It must present wage disparities by regions, cities, industries, education, occupation, and gender in a clear and intuitive manner.
- **Ensuring interpretability.** Some economic terminologies might not be familiar to all users. There should be clear labels and, if necessary, brief explanations of terms, to help users interpret the data without oversimplification.
- **Ensuring interactivity.** Users must be allowed to view the visualizations based on the required level of detail. This necessitates filters and interactive elements to compare trends intuitively.

## Visualization Problem

Wage disparities across the Philippine regions, industries, education levels, occupational status, and gender are crucial for policymakers, researchers, and the general public. Policymakers must go through published research papers or process the data to make informed decisions for labor and economic policies. An interactive dashboard could easily show the key indicators by region

and industry and provide heat maps highlighting the disparities. Researchers must process detailed data in spreadsheets to analyze these labor market trends. This project could provide dashboards filtered by region, urban-rural areas, sector, education, occupation, and gender. Lastly, the general public needs intuitive and accessible representations of wage trends. This project will provide infographics and interactive maps that are easily accessible and digestible to the public. Overall, this interactive visualization application can enable these stakeholders to explore wage trends and disparities easily and more efficiently. But this easy-to-navigate dashboard makes these data easily accessible without having the stakeholders process it themselves or read research papers.

## **Dataset Description**

### **Brief description**

The [Labor Force Survey \(LFS\) 2023](#) is a nationwide survey conducted by the **Philippine Statistics Authority (PSA)** to collect data on employment, unemployment, and underemployment. It provides key insights for labor market policies and economic planning.

**Source of Data:** Philippine Statistics Authority (PSA) – collected through household interviews on a quarterly and monthly basis

### **Collection Process**

- Conducted via household interviews covering all regions in the Philippines.
- Targets individuals aged 15 years and older, excluding Overseas Filipino Workers (OFWs).
- Uses a multi-stage sampling design to ensure representativeness.
- For instance, The November 2022 survey covered 10,965 sample households (sample size may vary in other rounds of surveys)

### **Dataset Structure**

- Each row represents individual respondents in the labor force
- Each column corresponds to various demographic, employment, and economic variables

## Key Variables

Variable	Name	Data Type	Unit	Notes
PUFREG	Region	Integer (converted into string)	From regions 1-17	Converted into Categories
PUFURB2015	Urban-Rural	string	Urban and rural	Converted into Categories
PUFC04_SEX	Sex	string	Male and female	Converted into Categories
PUFC07_GRADE	Highest Grade Completed	string (converted into integer to make categories)	Grade completed	Converted into Categories
PUFC16_PKB	Kind of Business (Primary Occupation and Sector)	string (converted into integer to make categories)	Primary occupation and under which sector they belong	Converted into Categories
PUFC25_PBASIC	Basic Pay Per Day (Primary Occupation)	String (converted to integer)	Peso wage per day	-

## Connection

Starting from 2022, the LFS has monthly releases all throughout the year. However, some variables do not exist across all months. In this case, the most important variable is basic pay per day (*PUFC25\_PBASIC*) as this visualization focuses on the daily wages across regions and across different geographic and demographic characteristics. With 12 available monthly dataset, only 4 have consistent data on this variable. However, only 3 data frames have all of the aforementioned variables. These are the months of January, April, and July LFS 2023, which would be concatenated and then be used for the visualization of this project.

The spatial aspect of this project would come into place using the regional variable (*PUFREG*) available across the three months. Using shape files from [Macoy Meija from GitHub](#) on the regional level of the Philippines, data of average basic pay across the regions can be visualized via a choropleth map.

### ***Initial Exploratory Data Analysis***

By ensuring that the data can be visualized, the average basic pay per day was calculated from different avenues: by regional, by urban and rural, by gender, by primary occupation, and educational attainment. The merged dataset is divided into these characteristics and average pay per day based on these characteristics were visualized. This serves as an experiment both on how the dataset needs to be cleaned and if the desired visualization is possible in the first place.

### ***Average Daily Pay Across Regions***

Below is the used code and visualization for average daily pay across regions via a bar graph. NCR has the highest average daily pay, while BARMM has the lowest.

```
#assigning words to urban and rural
urban_rural = {1: 'Urban', 2: 'Rural'}
LFS_merged['PUFURB2015'] = LFS_merged['PUFURB2015'].map(urban_rural)

# Average basic pay per day across regions
region_avg = LFS_merged.groupby('PUFREG')['PUFC25_PBASIC'].mean().reset_index()

# Converting region to string
region_avg['PUFREG'] = region_avg['PUFREG'].astype(str)

# Difference between Urban and Rural basic pay (for curiosity)
urban_rural_avg = LFS_merged.groupby('PUFURB2015')['PUFC25_PBASIC'].mean()
urban_rural_diff = urban_rural_avg['Urban'] - urban_rural_avg['Rural']
```

```

print("Average Basic Pay by Region:\n", region_avg)
print("\nDifference between Urban and Rural:", urban_rural_diff)

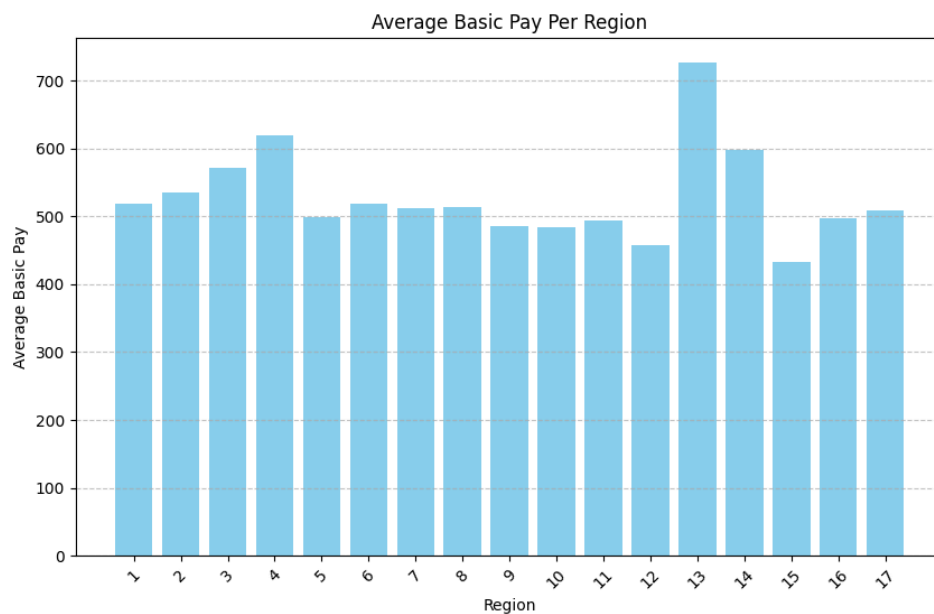
plt.figure(figsize=(10, 6))

# Plotting everything
plt.bar(region_avg['PUFREG'], region_avg['PUFC25_PBASIC'], color='skyblue')

plt.title("Average Basic Pay Per Region")
plt.xlabel("Region")
plt.ylabel("Average Basic Pay")
plt.xticks(rotation=45) # Rotate labels for better readability
plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.show()

```



### ***Testing Average Pay Across Regions, visualizing via a map***

Below is the code and first run of the visualization with a choropleth map. Three regions are suddenly missing. After trying to resolve the issue, the notebook in use suddenly started crashing and no further improvements were made. However, with this, it is possible to visualize the data via a map.

```
region_avg = LFS_merged.groupby("PUFREG",
as_index=False)["PUFC25_PBASIC"].mean()

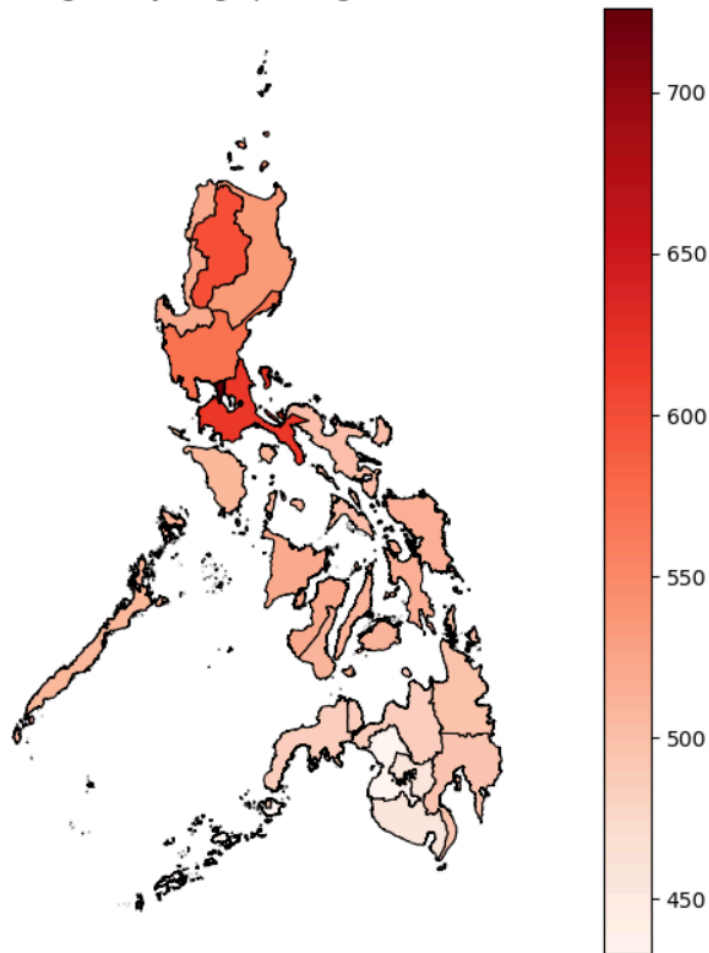
# Merge the shapefile with the wage data
LFS_merged = shapefile.merge(region_avg, left_on="ADM1_EN",
right_on="PUFREG", how="left")

# Convert to GeoDataFrame
LFS_merged = gpd.GeoDataFrame(LFS_merged, geometry="geometry")

# Plot the choropleth
fig, ax = plt.subplots(figsize=(12, 8))
LFS_merged.plot(column="PUFC25_PBASIC", cmap="Reds", linewidth=0.8,
edgecolor="black", legend=True, ax=ax)

plt.title("Average Daily Wage per Region")
plt.axis("off")
plt.show()
```

Average Daily Wage per Region



### *Average Daily Pay Across Regions, between Urban and Rural*

Below is the used code and visualization for average daily pay between Urban and Rural across regions via a bar graph. Urban areas generally have higher wages in comparison to rural areas.

```
# Region with Urban Rural
```

```
urban_rural_region_avg = LFS_merged.groupby(['PUFREG',  
'PUFURB2015'])['PUFC25_PBASIC'].mean().unstack()
```

```
# Calculate the difference between Urban and Rural per region (for curiosity)
```

```
urban_rural_region_avg['Pay_Difference'] = urban_rural_region_avg['Urban'] -
```

```
urban_rural_region_avg['Rural']
```

```
print(urban_rural_region_avg)
```

```
urban_rural_region_avg[['Urban', 'Rural']].plot(kind='bar', figsize=(10, 6), colormap='plasma')
```

```
#Plotting stuff
```

```
plt.title("Urban vs. Rural Basic Pay Per Region")
```

```
plt.xlabel("Region")
```

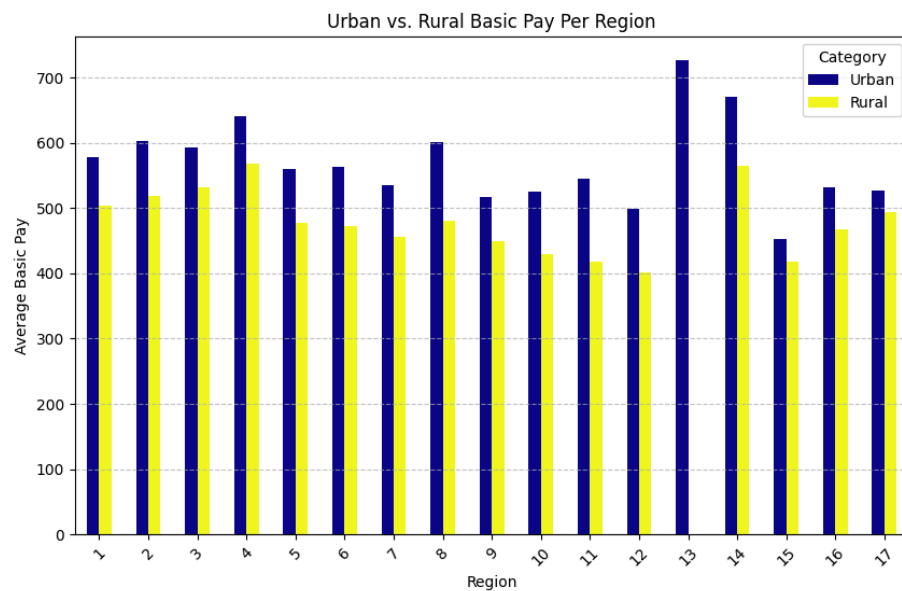
```
plt.ylabel("Average Basic Pay")
```

```
plt.xticks(rotation=45)
```

```
plt.legend(title="Category")
```

```
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

```
plt.show()
```





### *Average Daily Pay Across Regions, between Sexes*

Below is the used code and visualization for average daily pay between Male and Female across regions via a bar graph. On average, females have higher basic pay than males across all the regions. However, there is only a marginal difference in NCR.

```
# Assigning categories for the gender variable
sex_mapping = {1: 'Male', 2: 'Female'}
LFS_merged['PUFC04_SEX'] = LFS_merged['PUFC04_SEX'].map(sex_mapping)

# Group by both Region and Sex, then calculate the mean basic pay
sex_region_avg = LFS_merged.groupby(['PUFREG',
'PUFC04_SEX'])['PUFC25_PBASIC'].mean().unstack()

# Calculate the difference between Male and Female per region
sex_region_avg['Pay_Difference'] = sex_region_avg['Male'] - sex_region_avg['Female']

# Display result
print(sex_region_avg)

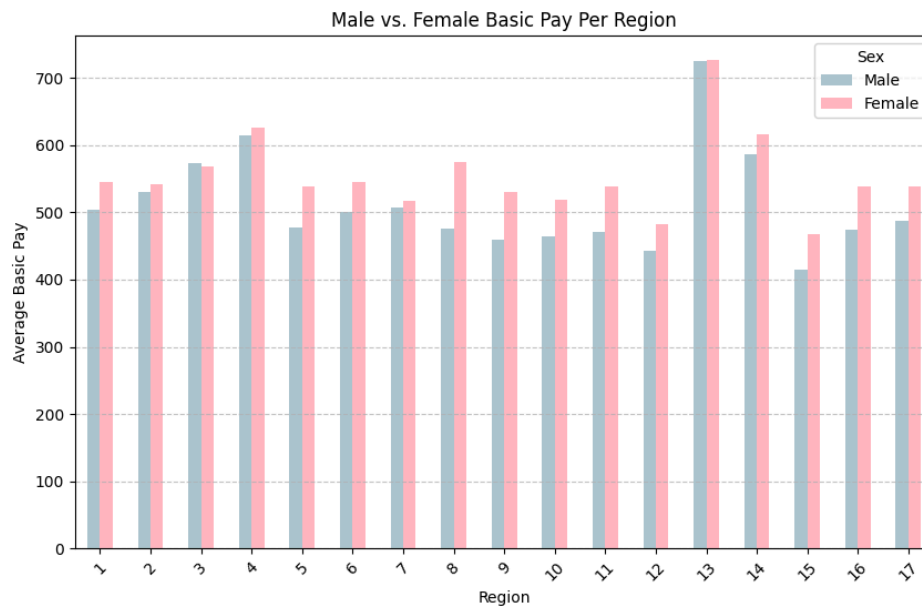
# Bar graph comparing Male vs. Female pay per region
colors = ['#AEC6CF', '#FFB6C1'] # Pastel blue for Male, Pastel pink for Female

# Bar graph comparing Male vs. Female pay per region
sex_region_avg[['Male', 'Female']].plot(kind='bar', figsize=(10, 6), color=colors)

plt.title("Male vs. Female Basic Pay Per Region")
plt.xlabel("Region")
plt.ylabel("Average Basic Pay")
plt.xticks(rotation=45)
plt.legend(title="Sex")
```

```
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

```
plt.show()
```



### ***Average Daily Pay Across Regions, Among Sectors***

Below is the used code and visualization for average daily pay among different sectors across regions via a bar graph. Services is the highest, followed by manufacturing, and lastly agriculture.

```
#primary occupation range string to float
```

```
LFS_merged['PUFC16_PKB'] = pd.to_numeric(LFS_merged['PUFC16_PKB'],
errors='coerce')
```

```
# PUFC25_PBASIC to integer
```

```
LFS_merged['PUFC25_PBASIC'] = pd.to_numeric(LFS_merged['PUFC25_PBASIC'],
errors='coerce')
```

```
#dropping NAs
```

```
LFS_merged = LFS_merged.dropna(subset=['PUFC16_PKB'])
```

```
# Define categorization function
```

```
def categorize(PUFC16_PKB):
```

```
    if PUFC16_PKB <= 3:
```

```
        return 'Agriculture'
```

```
    elif 4 <= PUFC16_PKB <= 43:
```

```
        return 'Industry'
```

```
    else:
```

```
        return 'Services'
```

```
LFS_merged['Occupation'] = LFS_merged['PUFC16_PKB'].apply(categorize)
```

```
# Group by both Region and PUFC16_Category, then calculate the mean basic pay
```

```
category_region_avg = LFS_merged.groupby(['PUFREG',  
'Occupation'])['PUFC25_PBASIC'].mean().unstack()
```

```
# difference between Agriculture and Services per region
```

```
category_region_avg['Pay_Difference'] = category_region_avg['Services'] -  
category_region_avg['Agriculture'] #add more if you are curious
```

```
# Display result
```

```
print(category_region_avg)
```

```
# Bar graph comparing PUFC16_Category pay per region
```

```
colors = ['#A2D9CE', '#F7DC6F', '#EC7063'] # Teal for Agriculture, Yellow for Industry, Red  
for Services
```

```
category_region_avg[['Agriculture', 'Industry', 'Services']].plot(kind='bar', figsize=(10, 6),
```

```
color=colors)
```

```
plt.title("Basic Pay Across Sectors Per Region")
```

```
plt.xlabel("Region")
```

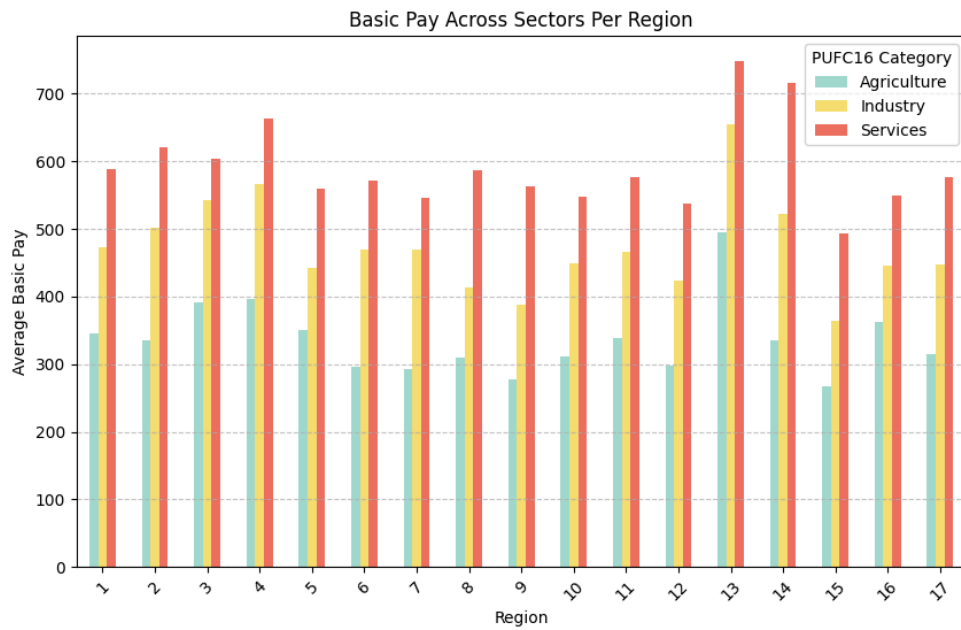
```
plt.ylabel("Average Basic Pay")
```

```
plt.xticks(rotation=45)
```

```
plt.legend(title="PUFC16 Category")
```

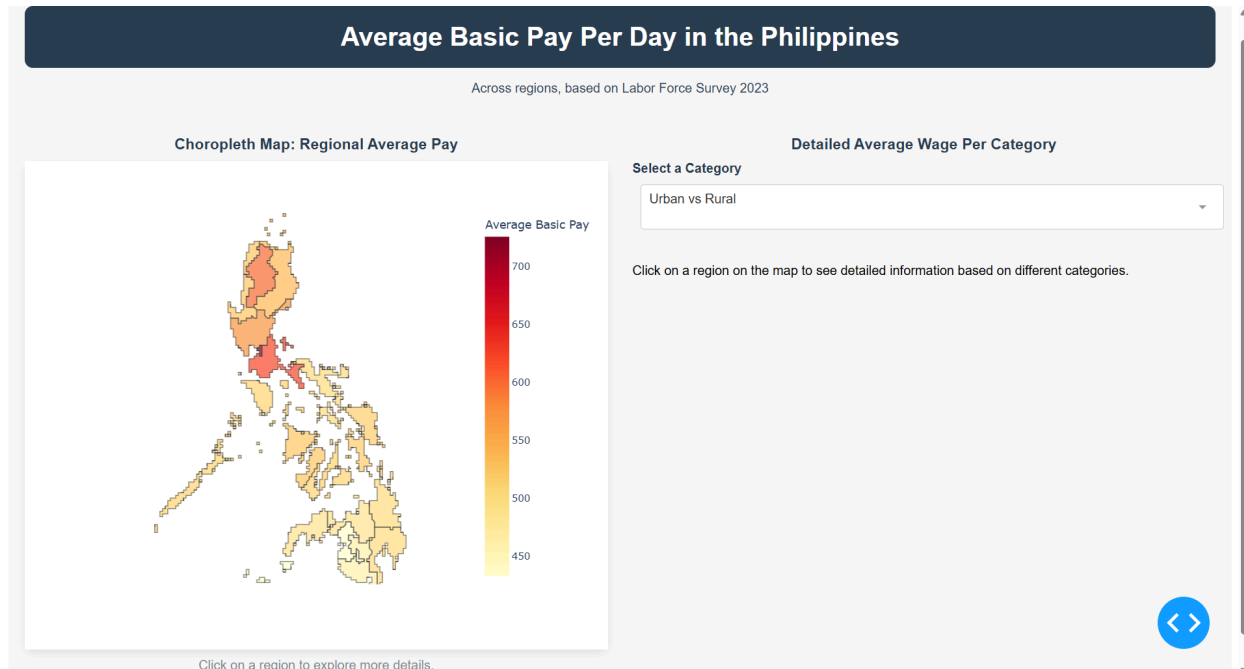
```
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

```
plt.show()
```



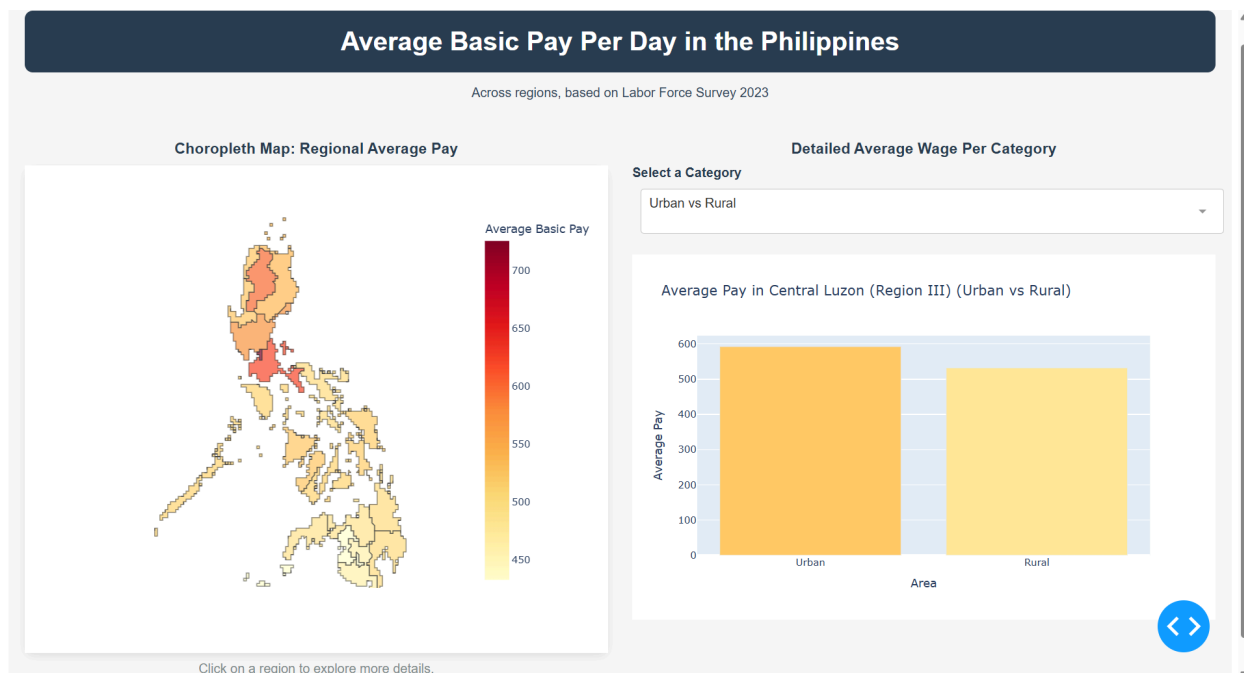
# 1 | Prototype Design

## 1.1 | General

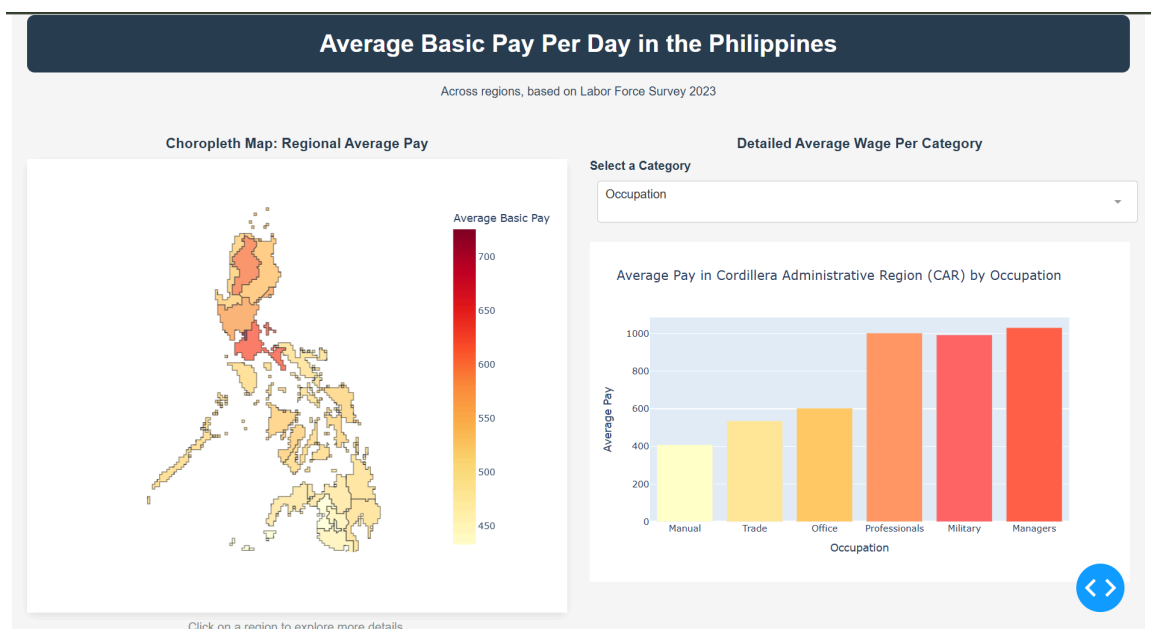


The dashboard provides a visual representation of wage data through a choropleth map and a summary of key variables through bar graphs. By default, it displays the average basic pay at the regional level via a choropleth map on the left and an empty “*Detailed Average Wage per Category*” with a dropdown on the right. Instructions saying “*Click on a region to explore more details*” and “*Click on a region on the map to see detailed information based on different categories*” can be seen to let users know that the map is clickable and the bar graphs would be dynamic, meaning it can change depending on the region clicked.

## 1.2 | Specific

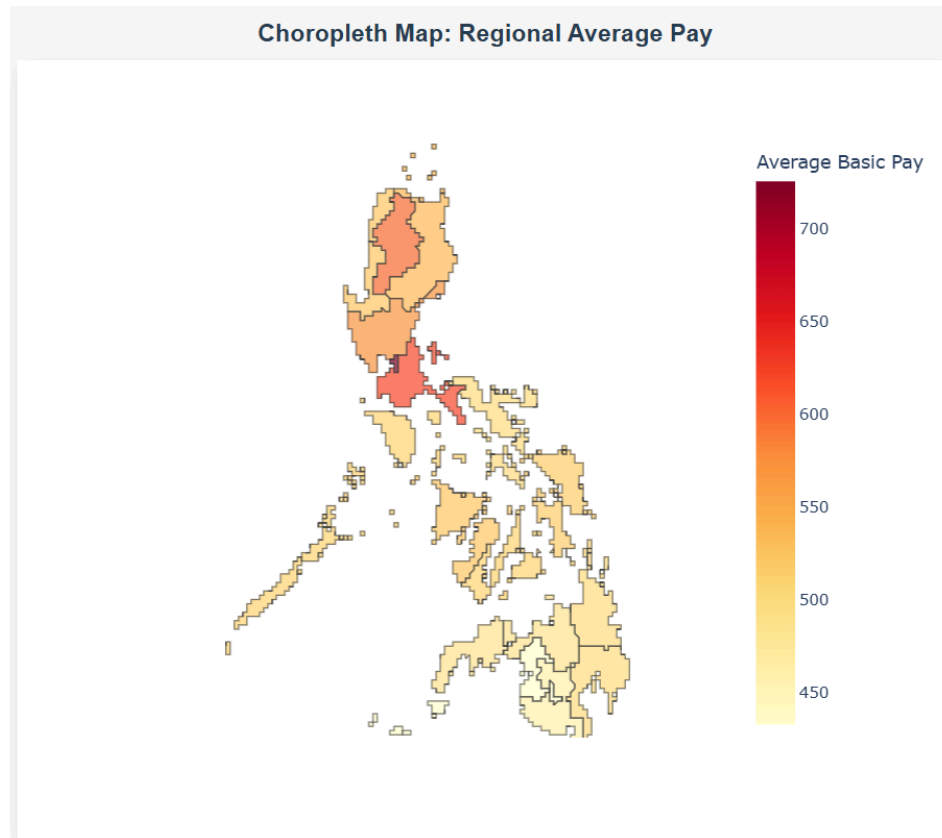


After clicking on a region, the left part of the dashboard will reveal a bar graph based on the selected category from the dropdown. Its default is set to Urban vs Rural, but can be changed accordingly based on what the user chooses to be the dropdown.



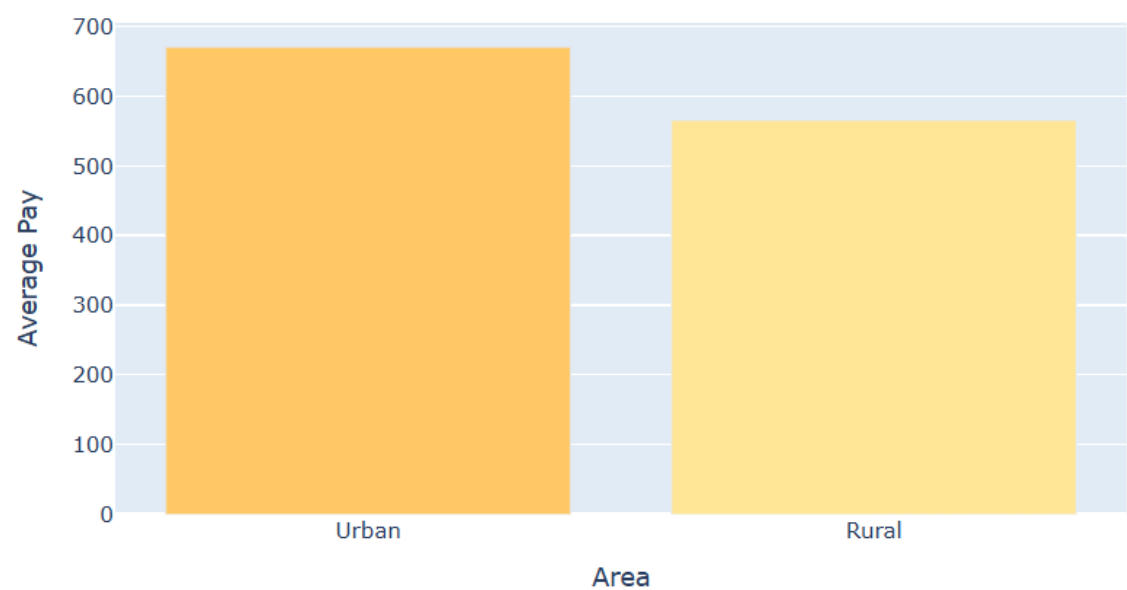
## 2 | Visualization Choices

### 2.1 | Choropleth Map



<b>Idiom</b>	Choropleth Map
<b>Data</b>	Average basic pay in the Philippines by region
<b>Channels</b>	<p>Color used to represent wage levels:</p> <ul style="list-style-type: none"><li>• A gradient of red (from light yellow to deep red)<ul style="list-style-type: none"><li>- To represent the variation in wage levels</li><li>- Lighter yellow would mean lower average basic pay, deeper red would mean higher average basic pay</li><li>- Shows the intensity among the distribution of the average basic pay</li></ul></li></ul>
<b>Task</b>	Discover regional wage disparities in the Philippines

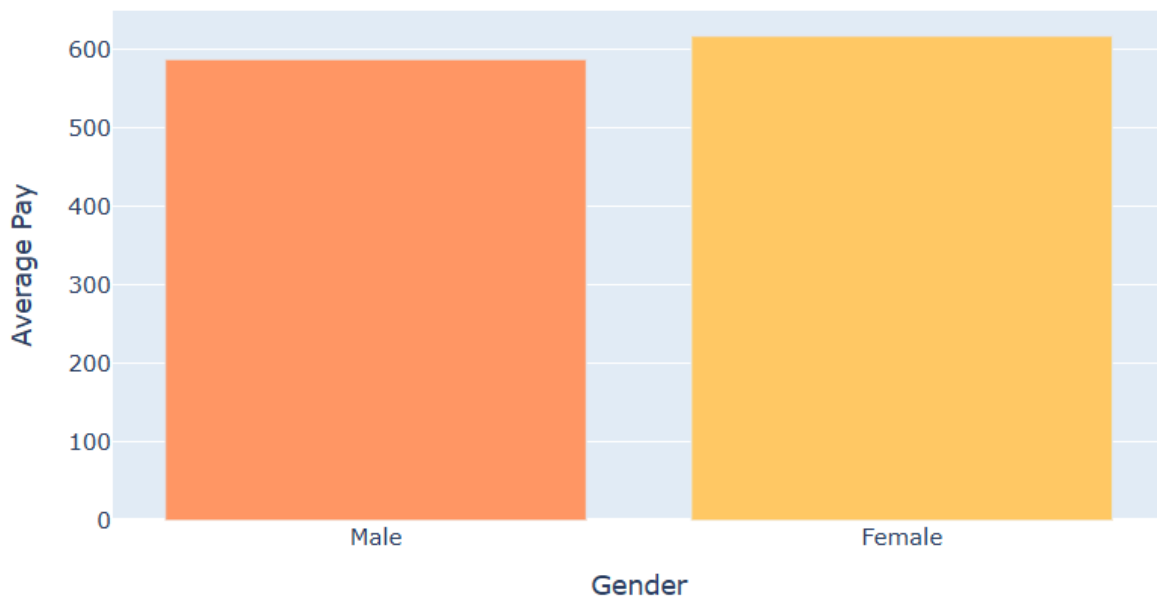
2.2 | Urban vs Rural



Idiom	Bar Chart
Data	On the x-axis: city types (urban and rural) On the y-axis: average basic pay
Channels	Colors used to differentiate categories: <ul style="list-style-type: none"><li>● Orange - Urban</li><li>● Yellow - Rural</li></ul>
Task	Compare the average basic pay between urban and rural areas

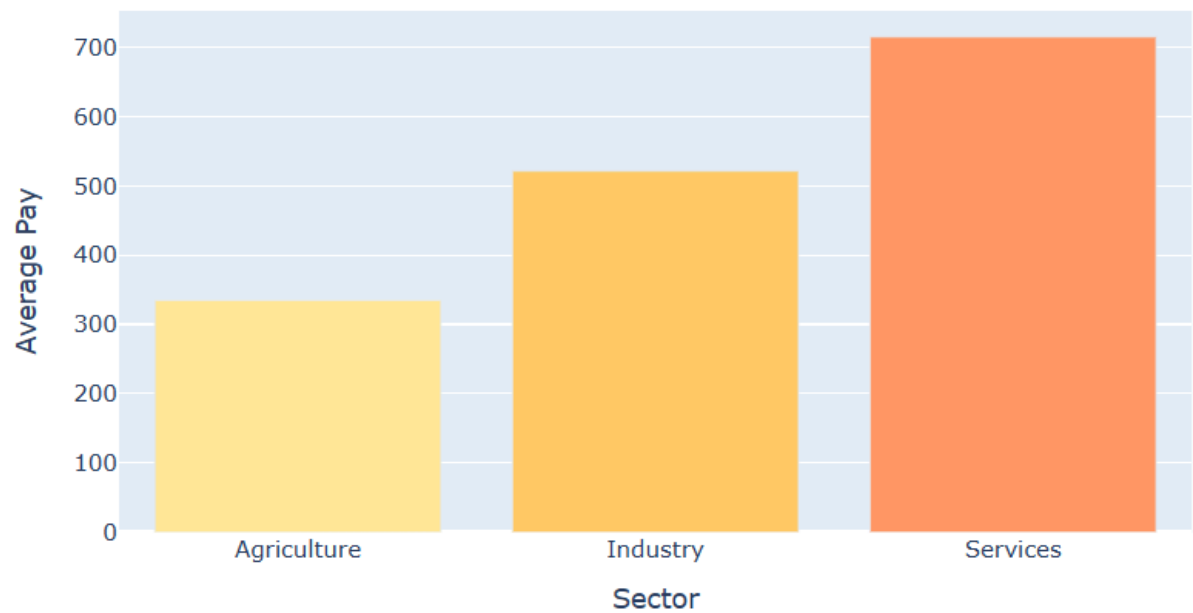


## 2.3 | Male vs Female



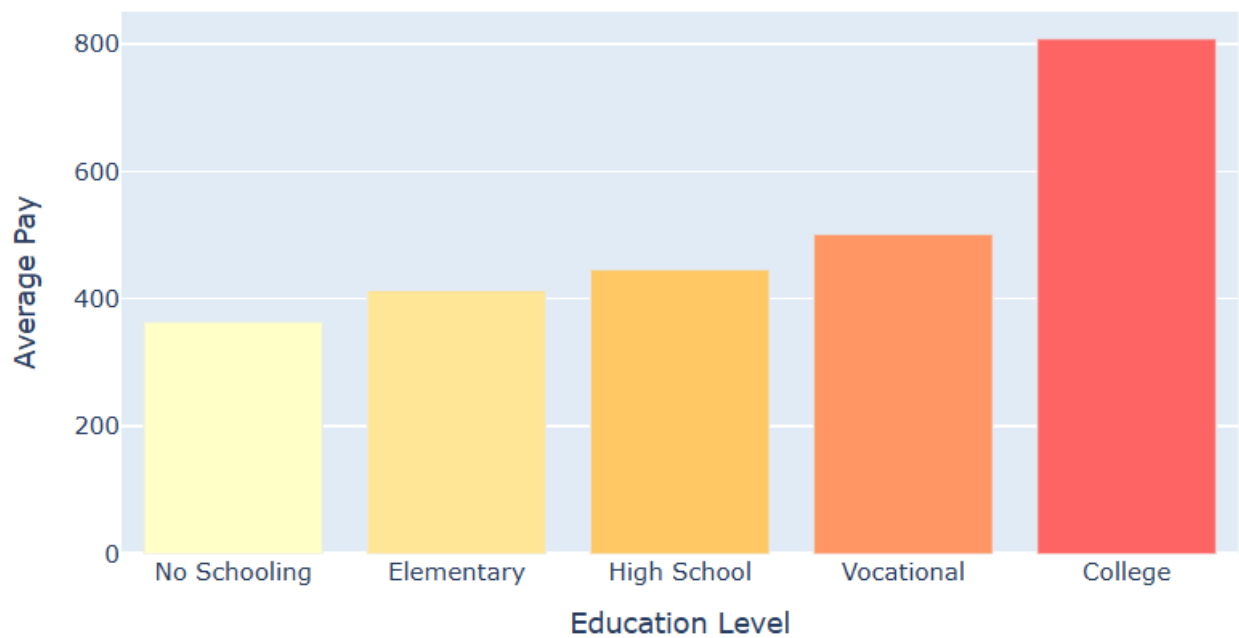
Idiom	Bar Chart
Data	On the x-axis: genders (male and female) On the y-axis: average basic pay
Channels	Colors used to differentiate categories: <ul style="list-style-type: none"><li>• Deep Orange - Male</li><li>• Deep Yellow - Female</li></ul>
Task	Compare the average basic pay between males and females

2.4 | Sector



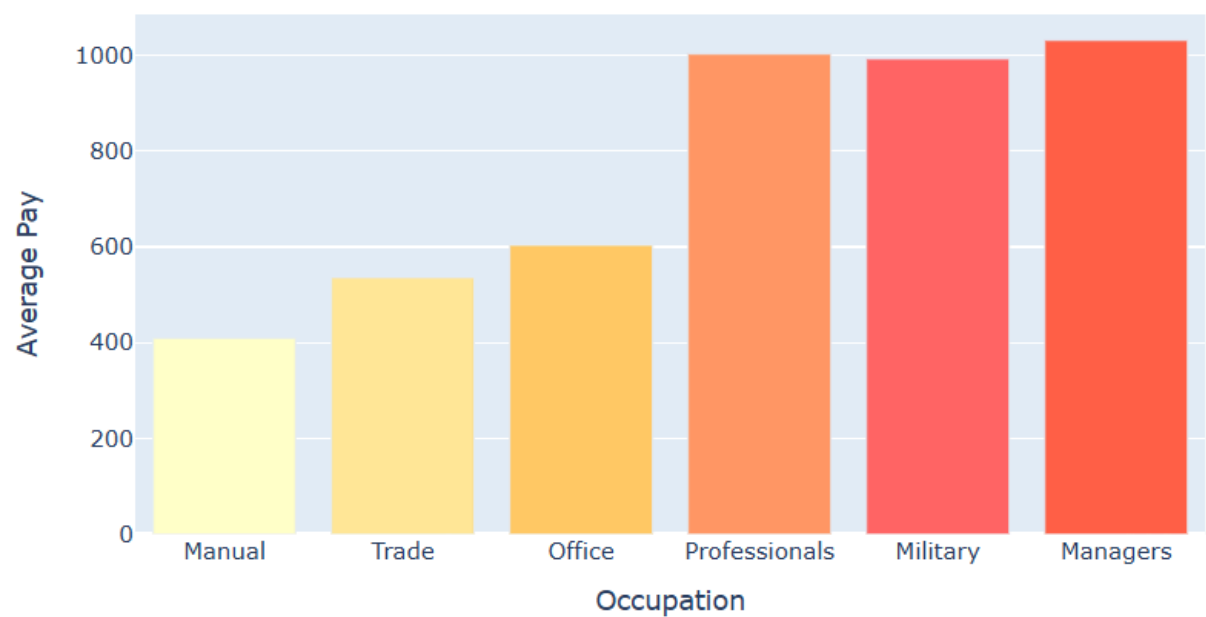
Idiom	Bar Chart
Data	On the x-axis: sectors (Agriculture, Industry, Services) On the y-axis: average basic pay
Channels	Colors used to differentiate categories: <ul style="list-style-type: none"><li>• Deep orange - Services</li><li>• Deep Yellow - Industry</li><li>• Light Yellow - Agriculture</li></ul>
Task	Compare the average basic pay among the sectors

2.5 | Education Level



Idiom	Bar Chart
Data	<div>On the x-axis: Highest educational attainment (Post-grad, College, High School, Elementary, No Schooling)</div> <div>On the y-axis: average basic pay</div>
Channels	Each Bar is in different shades of red, forming a gradient from light yellow to dark red, representing different education levels
Task	Compare the average basic pay among the different educational levels

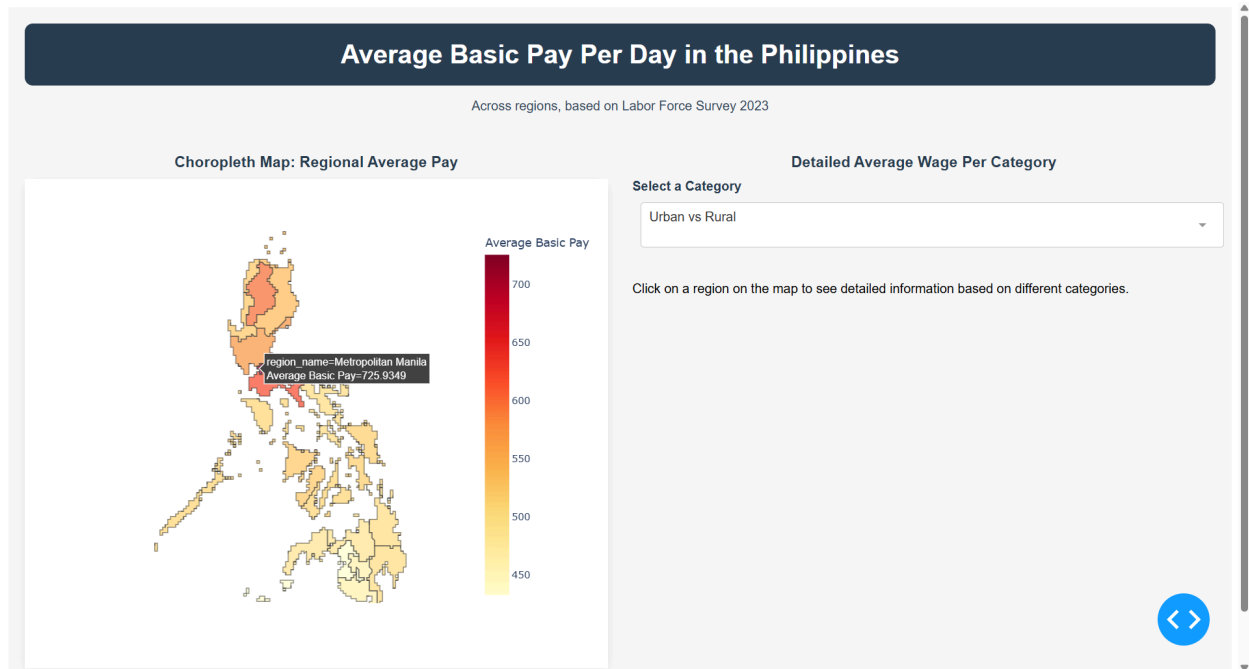
2.6 | Jobs



Idiom	Bar Chart
Data	On the x-axis: Aggregated general occupational groups On the y-axis: average basic pay
Channels	Each Bar is in different shades of red, forming a gradient from light yellow to dark red, representing different occupations
Task	Compare the average basic pay among the different occupational categories

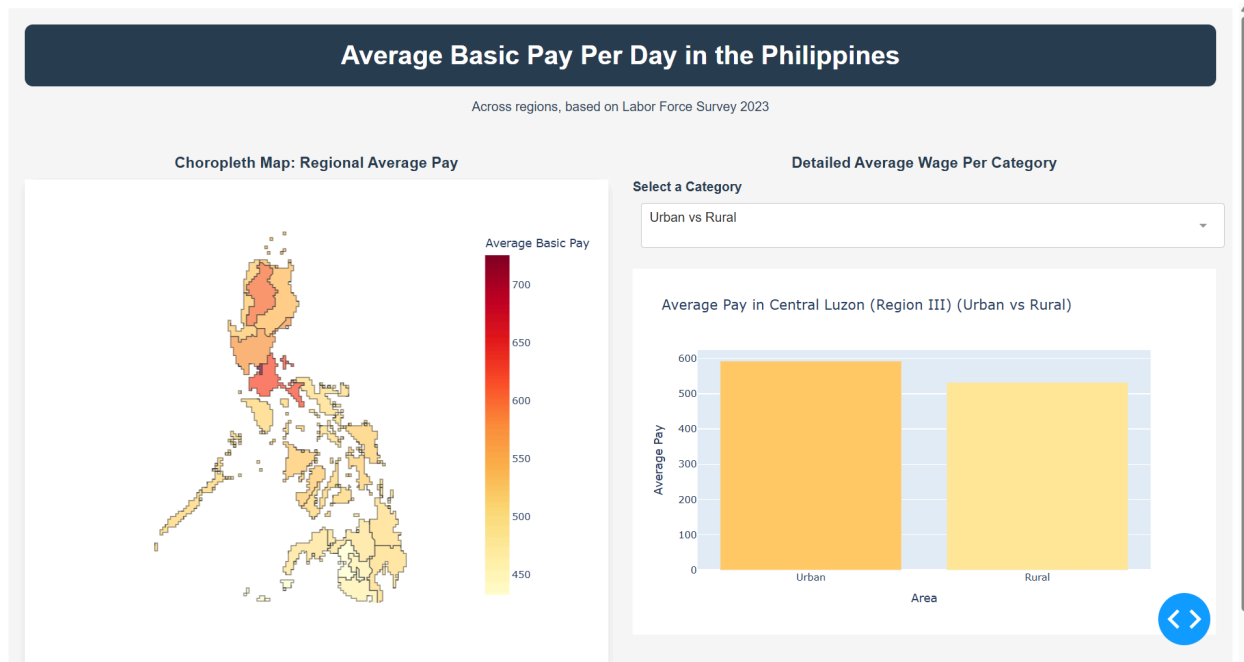
## 3 | Interactivity Techniques

### 3.1 | Hover



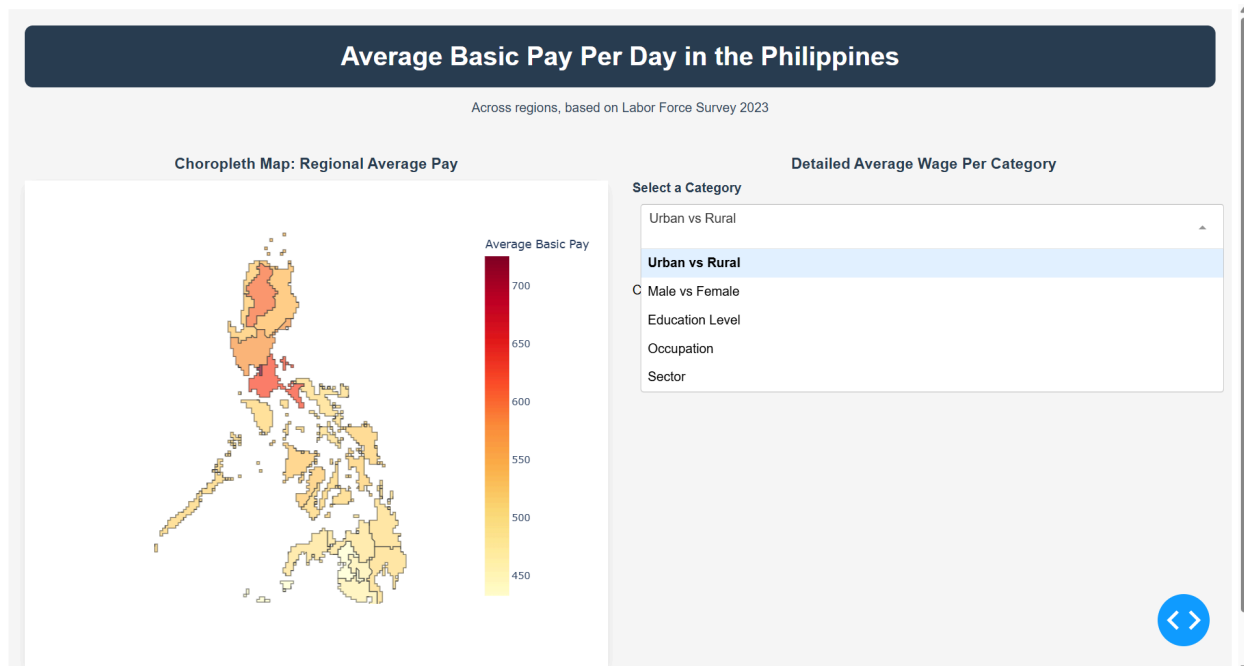
Hovering over an area on the map displays the name of the region or island and its exact average basic pay. This allows the user to discover the exact wage disparity between the areas of interest, as the colors only provide the direction of the disparity and not actual numbers.

### 3.2 | Clickable map



Users can explore region-specific wage data by clicking on areas in the choropleth map. This updates the "Detailed Average Wage Per Category" section to display bar charts comparison by different demographic categories such as urban-rural, gender, sector, education, and occupation. By clicking on a different region in the map, the bar graphs would change accordingly based on the information from that region.

### 3.3 | Dropdown



As mentioned earlier, the users can choose to compare data at different demographic categories, allowing them to adjust the details based on their needs. The dropdown selection updates the bar graphs based on what category is in the box. This interactive feature provides a more detailed preview, allowing users to compare wages across different characteristics with ease.