Malaysia Labour Force Analysis (2010–2025)
Prepared by
Name : Hanis Ilya binti Mohamad Shaharizad
Data Source: Department of Statistics Malaysia (DOSM)

1. Objective

This research looks at Malaysia's labour force, employment, and unemployment trends from 2010 to 2025. Using DOSM official data, the goal is to detect important changes in labor participation and the impact of economic events like as the COVID-19 epidemic.

2. Tools Used
   ● Python (Google Colab)
   ● pandas, matplotlib, seaborn
   ● Dataset: Monthly Principal Labour Force Statistics

3. Data Overview

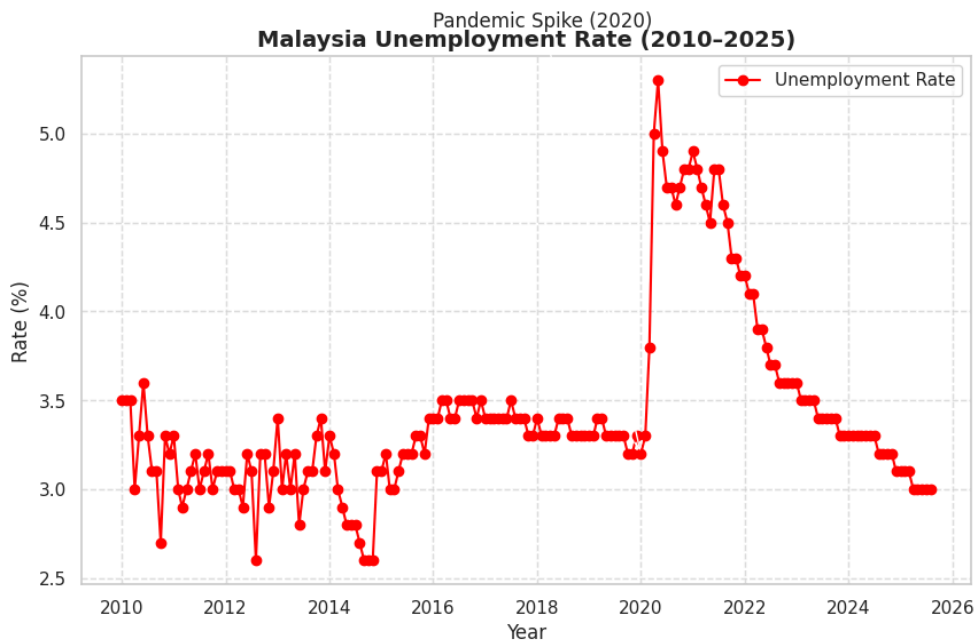| Column	Description
date	Month and year (2010–2025)
lf	Total labour force
lf_employed	Total employed persons
lf_unemployed	Total unemployed persons
lf_outside	Population outside labour force
p_rate	Labour force participation rate (%)
ep_ratio	Employment-to-population ratio (%)
u_rate	Unemployment rate (%)

4. Visualization

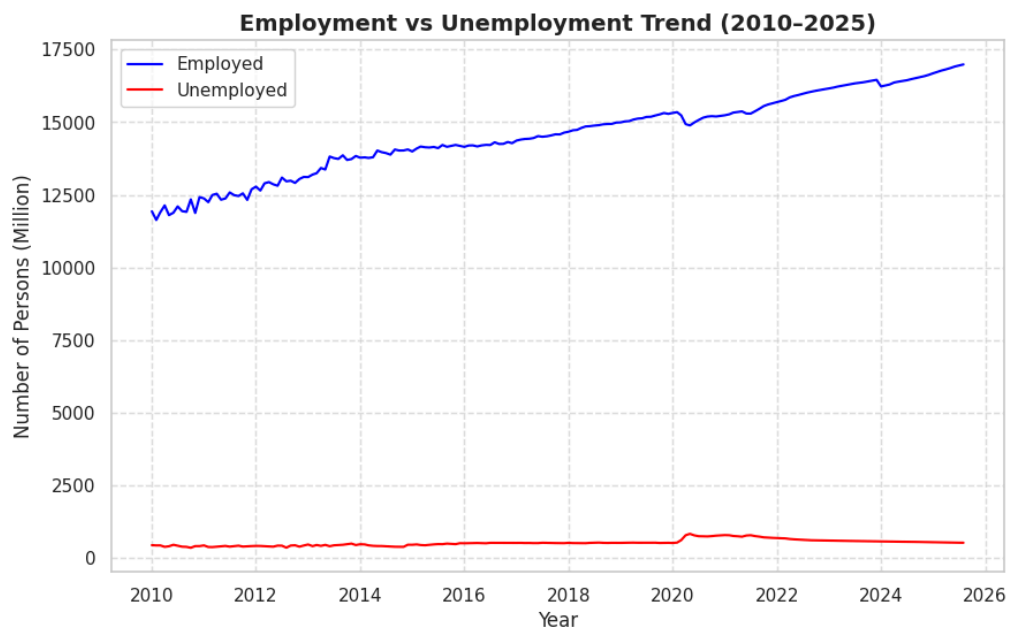**Figure 1:** Unemployment Rate (2010–2025)

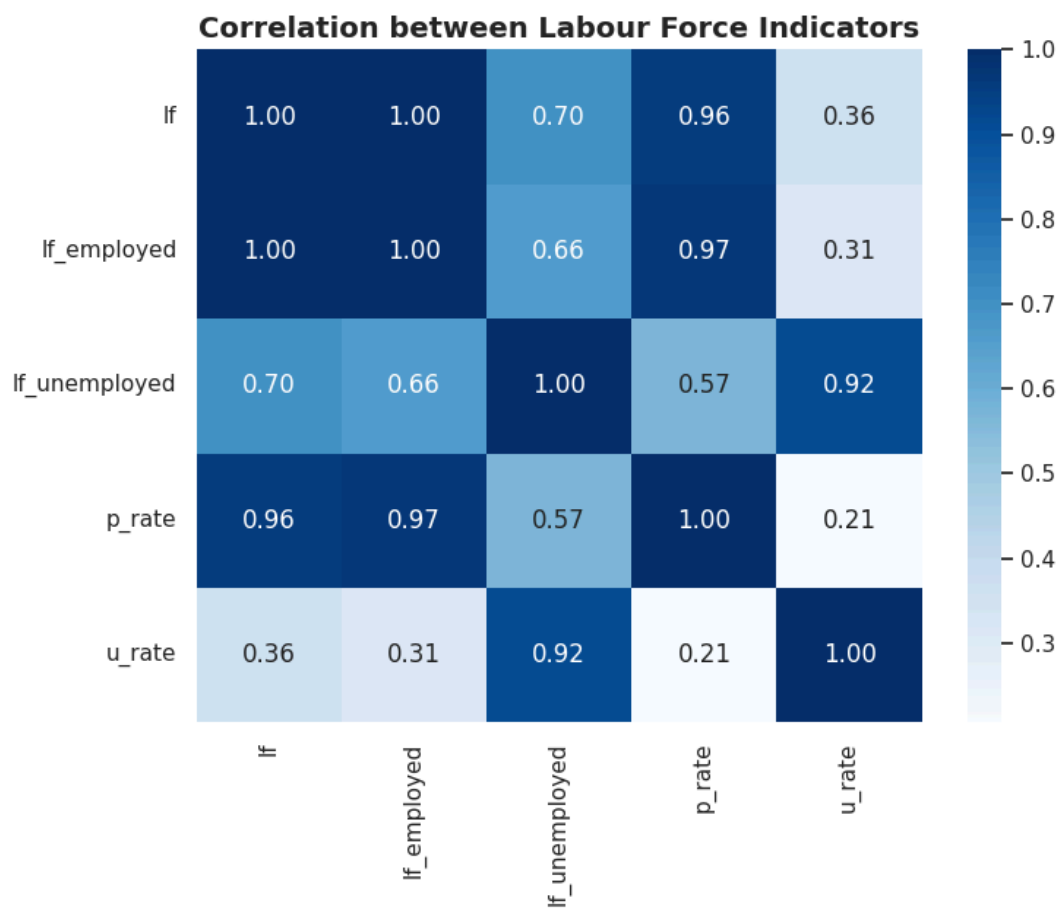**Figure 2:** Employment vs Unemployment Trend

Figure 3: Correlation between Labour Force Indicators

5. Key Findings

Unemployment Rate (2010–2025)
Malaysia's unemployment rate stayed very steady between 2010 and 2019, averaging approximately 3.3%. However, a large increase happened in 2020 due to the COVID-19 pandemic, when lockdowns and company closures resulted in major employment losses. The rate peaked at more than 4.5% before steadily falling in the years that followed as the economy reopened and recovery strategies were implemented. By 2025, unemployment had recovered to pre-pandemic levels, indicating better labor market conditions and economic resiliency.

Employment vs Unemployment Trend
The trend comparison of employment and unemployment shows an opposite trend: as employment climbed, unemployment decreased. The 2020-2021 period indicates a strong drop in employment and an increase in unemployment, consistent with pandemic disruptions. Employment has steadily rebounded since 2022, indicating substantial post-pandemic job growth and economic adaptation to the new labor environment, notably in the virtual and hybrid industries.

Correlation of Labour Force Indicators
The correlation heatmap shows significant positive correlations between the labor force, employment, and participation rate, implying that as more individuals enter the labour market, employment levels grow. Unemployment, on the other hand, has a slightly negative connection with employment and participation, implying that greater labour market activity leads to decreased unemployment. These relationships reflect an effective and consistent labor market structure in which increased involvement translates efficiently into opportunities for employment.

6. Conclusion

Overall, from 2010 to 2025, Malaysia's labor market had a steady pattern of recovery and growth. Although the COVID-19 epidemic resulted in a brief increase in unemployment and a decline in employment, the economy recovered significantly in the following years. Rising labour force participation and a constant employment-to-population ratio suggest increased worker engagement and job availability. By 2025, Malaysia's unemployment rate will have reverted to pre-pandemic levels, demonstrating good economic policy, worker adaptation, and long-term national resilience.

Documentation

Step 1: Data Loading and Cleaning

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Display settings
pd.set_option('display.max_columns', None)
sns.set(style="whitegrid")
```

```python
from google.colab import filess
uploaded = files.upload()
```

Choose Files No file chosen   Upload widget is only available when the cell has been executed in the current browse
Saving monthly_principal_labour_force_statistics.csv to monthly_principal_labour_force_statistics.csv

```python
import pandas as pd

df = pd.read_csv("monthly_principal_labour_force_statistics.csv")
df.head()
```

	date	lf	lf_employed	lf_unemployed	lf_outside	p_rate	ep_ratio	u_rate
0	2010-01-01	12367.9	11931.2	436.7	6927.5	64.1	61.8	3.5
1	2010-02-01	12059.8	11632.3	427.4	7225.1	62.5	60.3	3.5

```python
df.info()
df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 188 entries, 0 to 187
Data columns (total 8 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   date           188 non-null    object
 1   lf             188 non-null    float64
 2   lf_employed    188 non-null    float64
 3   lf_unemployed  188 non-null    float64
 4   lf_outside     188 non-null    float64
 5   p_rate         188 non-null    float64
 6   ep_ratio       188 non-null    float64
 7   u_rate         188 non-null    float64
dtypes: float64(7), object(1)
memory usage: 11.9+ KB
```

Step 2: Exploratory Data Analysis (EDA)

```
df.info()
df.describe()
df.head()


<class 'pandas.core.frame.DataFrame'>
RangeIndex: 188 entries, 0 to 187
Data columns (total 8 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   date           188 non-null    object
 1   lf             188 non-null    float64
 2   lf_employed    188 non-null    float64
 3   lf_unemployed  188 non-null    float64
 4   lf_outside     188 non-null    float64
 5   p_rate         188 non-null    float64
 6   ep_ratio       188 non-null    float64
 7   u_rate         188 non-null    float64
dtypes: float64(7), object(1)
memory usage: 11.9+ KB
```

	date	lf	lf_employed	lf_unemployed	lf_outside	p_rate	ep_ratio	u_rate
0	2010-01-01	12367.9	11931.2	436.7	6927.5	64.1	61.8	3.5

Step 3: Visualization

```python
import matplotlib.pyplot as plt

plt.figure(figsize=(10,6))
plt.plot(df['date'], df['u_rate'], color='red', marker='o', label='Unemployment Rate')
plt.title("Malaysia Unemployment Rate (2010-2025)", fontsize=14, fontweight='bold')
plt.xlabel("Year")
plt.ylabel("Rate (%)")
plt.grid(True, linestyle='--', alpha=0.6)

# Convert date to datetime if not already
df['date'] = pd.to_datetime(df['date'])

# Find the data point closest to 2020
point_2020 = df[df['date'].dt.year == 2020].iloc[0]

# Annotate the 2020 spike
plt.annotate('Pandemic Spike (2020)',
             xy=(point_2020['date'], point_2020['u_rate']),
             xytext=(pd.Timestamp('2016-01-01'), df['u_rate'].max()+0.3),
             arrowprops=dict(facecolor='black', arrowstyle='->'))

plt.legend()
plt.show()
```

```python
plt.figure(figsize=(10,6))
plt.plot(df['date'], df['lf_employed'], color='blue', label='Employed')
plt.plot(df['date'], df['lf_unemployed'], color='red', label='Unemployed')
plt.title("Employment vs Unemployment Trend (2010-2025)", fontsize=14, fontweight='bold')
plt.xlabel("Year")
plt.ylabel("Number of Persons (Million)")
plt.grid(True, linestyle='--', alpha=0.6)
plt.legend()
plt.show()
```

```python
import seaborn as sns

plt.figure(figsize=(8,6))
sns.heatmap(df[['lf','lf_employed','lf_unemployed','p_rate','u_rate']].corr(),
            annot=True, cmap='Blues', fmt=".2f")
plt.title("Correlation between Labour Force Indicators", fontsize=14, fontweight='bold')
plt.show()
```