Mini Project 2
IST652 - Scripting for Data Analysis
Name: Isha Nitin Agrawal
SUID: 635393581

Comparative Analysis of the 2020 and 2024 US Presidential Elections.

Goal:

The analysis focuses on comparing the 2020 and 2024 US Presidential Elections to identify shifts in voter preferences, changes in vote margins, and trends across swing states, strongholds, and overall electoral outcomes. By analyzing changes in vote percentages, electoral votes, and vote margins for both major candidates, this study highlights key shifts in political dynamics and voter behavior over the two election cycles.

The Data and Its Source:

The dataset for this analysis compares the US Presidential Election results from 2020 and 2024. The data was gathered from the following sources:

1. **2020 Data**: The 2020 election data was obtained from The American Presidency Project. It contains detailed information about vote counts, percentages, and electoral votes for Joseph R. Biden and Donald Trump.
2. **2024 Data**: The 2024 election data was also obtained from The American Presidency Project but the data was incomplete and had many missing values which were important for the analysis, hence I filled the important data from CNN Elections.

   To ensure consistency between the two datasets and enable comparison, I created a **common column named State_Code**, which serves as a unique identifier for states in both datasets.

Dataset Overview:

2020 Dataset:

Columns:

STATE: Name of the state.

Joseph R. Biden Votes: Total votes received by Joseph R. Biden.

Joseph R. Biden %: Percentage of votes received by Joseph R. Biden.

Joseph R. Biden EV: Electoral votes secured by Joseph R. Biden.

Donald Trump Votes: Total votes received by Donald Trump.

Donald Trump %: Percentage of votes received by Donald Trump.

Donald Trump EV: Electoral votes secured by Donald Trump.

State_Code: Unique state identifier (e.g., AL for Alabama).

	STATE	Joseph R. Biden Votes	Joseph R. Biden %	Joseph R. Biden EV	Donald Trump Votes	Donald Trump %	Donald Trump EV	State_Code
1	Alabama	849624.0	36.57		1441170.0	62.03	9.0	AL
2	Alaska	153778.0	42.77		189951.0	52.83	3.0	AK
3	Arizona	1672143.0	49.36	11.0	1661686.0	49.06		AZ
4	Arkansas	423932.0	34.78		760647.0	62.4	6.0	AR
5	California	11110250.0	63.48	55.0	6006429.0	34.32		CA

2024 Dataset:

Columns:

STATE: Name of the state.

Kamala Harris Votes: Total votes received by Kamala Harris.

Kamala Harris %: Percentage of votes received by Kamala Harris.

Kamala Harris EV: Electoral votes secured by Kamala Harris.

Donald Trump Votes: Total votes received by Donald Trump.

Donald Trump %: Percentage of votes received by Donald Trump.

Donald Trump EV: Electoral votes secured by Donald Trump.

State_Code: Unique state identifier (e.g., AL for Alabama).

	STATE	Kamala Harris Votes	Kamala Harris %	Kamala Harris EV	Donald Trump Votes	Donald Trump %	Donald Trump EV	State_Code
1	Alabama	769391.0	34.1	0	1457704.0	64.6	9.0	AL
2	Alaska	139812.0	41.4	0	184204.0	54.5	3.0	AK
3	Arizona	1582860.0	46.7	0	1770242.0	52.2	11.0	AZ
4	Arkansas	396905.0	33.5	0	759241.0	64.2	6.0	AR
5	California	9183800.0	58.6	54	5988823.0	38.2	0.0	CA

Data Exploration and Cleaning Steps:

1. **Missing Data:**
   - The **Electoral Votes (EV)** column had missing values where a candidate had not won the state. These were filled with 0 to standardize the data, as only the winning candidate in each state is allocated EVs.
2. **Common Identifier:**
   - The State_Code column was added to both datasets to enable merging based on state information.
3. **Validation:**
   - The datasets were inspected to ensure consistency across state names and codes.
   - Columns for both datasets were standardized to ensure compatibility during analysis.
4. **Handling Null Values:**
   - Null values for votes or percentages were replaced with 0 to ensure smooth calculations and accurate comparisons.
5. **Generated Data for 2024:**
   - While the 2020 dataset was directly sourced, the 2024 dataset had to be generated based on available projections from CNN. This required additional validation and formatting to align with the 2020 dataset structure.

```
#Filling the null values with 0
data_2020["Joseph R. Biden Votes"].fillna(0, inplace=True)
data_2020["Joseph R. Biden %"].fillna(0, inplace=True)
data_2020["Joseph R. Biden EV"].fillna(0, inplace=True)
data_2020["Donald Trump Votes"].fillna(0, inplace=True)
data_2020["Donald Trump %"].fillna(0, inplace=True)
data_2020["Donald Trump EV"].fillna(0, inplace=True)

data_2024["Kamala Harris Votes"].fillna(0, inplace=True)
data_2024["Kamala Harris %"].fillna(0, inplace=True)
data_2024["Kamala Harris EV"].fillna(0, inplace=True)
data_2024["Donald Trump Votes"].fillna(0, inplace=True)
data_2024["Donald Trump %"].fillna(0, inplace=True)
data_2024["Donald Trump EV"].fillna(0, inplace=True)

## Standardize column names
data_2020.rename(columns={"Joseph R. Biden Votes": "Biden Votes","Joseph R. Biden %": "Biden %","Joseph R. Biden EV": "Biden EV", "Donald Trump Votes": "Trump Vot
data_2024.rename(columns={"Kamala Harris Votes": "Kamala Votes", "Kamala Harris %": "Kamala %", "Kamala Harris EV": "Kamala EV","Donald Trump Votes": "Trump Votes'


print(data_2020.head())

print(data_2024.head())


#making  a copy of the data
copy_data_2020 = data_2020.copy()
copy_data_2024 = data_2024.copy()
```

```
       STATE  Biden Votes  Biden %  Biden EV  Trump Votes  Trump %  Trump EV  \
0    Alabama     849624.0    36.57       0.0    1441170.0    62.03       9.0
1     Alaska     153778.0    42.77       0.0     189951.0    52.83       3.0
2    Arizona    1672143.0    49.36      11.0    1661686.0    49.06       0.0
3   Arkansas     423932.0    34.78       0.0     760647.0    62.40       6.0
4  California   11110250.0    63.48      55.0    6006429.0    34.32       0.0

  State_Code
0         AL
1         AK
2         AZ
3         AR
4         CA
       STATE  Kamala Votes  Kamala %  Kamala EV  Trump Votes  Trump %  \
0    Alabama      769391.0      34.1          0    1457704.0     64.6
1     Alaska      139812.0      41.4          0     184204.0     54.5
2    Arizona     1582860.0      46.7          0    1770242.0     52.2
3   Arkansas      396905.0      33.5          0     759241.0     64.2
```

Three Comparison Questions with Units of Analysis and Computation Methods:

1. Summarizing Information from the 2024 Presidential Election

- **Unit of Analysis**: Total votes, average percentage of votes, and Electoral Votes (EVs) for each candidate.

- **Computation**: The 2024 dataset was analyzed to compute the total votes, average percentages, and EVs secured by Kamala Harris and Donald Trump.
    - Kamala Harris received **73.85M votes** (43.50% average) with **266 EVs**.
    - Donald Trump secured **76.94M votes** (50.61% average) with **253 EVs**.
- **Purpose**: This provides an overview of each candidate's performance in the 2024 elections, highlighting key differences in vote distribution and EV allocation.

```
# Summarize total votes and percentage for each candidate
summary_2024 = {
    "Total Votes (Kamala)": copy_data_2024["Kamala Votes"].sum(),
    "Average % (Kamala)": copy_data_2024["Kamala %"].mean(),
    "Total Electoral Votes (Kamala)": copy_data_2024["Kamala EV"].sum(),
    "Total Votes (Trump)": copy_data_2024["Trump Votes"].sum(),
    "Average % (Trump)": copy_data_2024["Trump %"].mean(),
    "Total Electoral Votes (Trump)": copy_data_2024["Trump EV"].sum(),
}

# Convert to DataFrame for display
summary_df_2024 = pd.DataFrame([summary_2024])
print("Summary of 2024 Election Data")
print(summary_df_2024)

#save the dataframe as a csv file
summary_df_2024.to_csv("summary_df_2024_Q1.csv")
```

```
Summary of 2024 Election Data
   Total Votes (Kamala)  Average % (Kamala)  Total Electoral Votes (Kamala)  \
0            73857695.0           43.501961                             266

   Total Votes (Trump)  Average % (Trump)  Total Electoral Votes (Trump)
0           76943544.0          50.619608                          253.0
```

Key Findings from this Q1 analysis:

1. We can see that in 2024 elections, **Kamala Harris** recieved **73857695** total votes and her average percentage was **43.50%**
2. On the other hand, we can see that **Donald Trump** has recieved **76943544** total votes and his average percentage was **50.61**
3. Additionally one interesting thing which we can notice is Total Electoral votes of kamala harris is 266 and that of donald trump is 253 which is less than kamala. This usually happens when:

    a) **Kamala Wins High-EV States:**

    Kamala's EVs are higher because she likely won states like California, New York, Illinois, and other Democratic-leaning large states with significant EV counts. These states contribute disproportionately to her EV total. Trump Wins Popular Vote in Strongholds:

    b) **Trump wins more total popular votes due to sweeping victories in Republican strongholds (e.g., Alabama, Texas, and other rural states).** These states often have fewer EVs relative to their population sizes. Overall Summary:

Kamala Harris: Secured 266 EVs by winning key states with high EV counts.

Donald Trump: Secured 253 EVs despite having more total popular votes because his victories were concentrated in smaller states with lower EVs

		Total Votes (Kamala)	Average % (Kamala)	Total Electoral Votes (Kamala)	Total Votes (Trump)	Average % (Trump)	Total Electoral Votes (Trump)
1	0	73857695.0	43.50196078431373	266	76943544.0	50.61960784313726	253.0

2. Categorizing States Based on Voting Patterns (2024 Data)

- **Unit of Analysis**: Categorization of states into **Kamala Strongholds**, **Trump Strongholds**, and **Swing States** based on percentage thresholds (>60% votes for strongholds, otherwise Swing States).
- **Computation**: Aggregated total votes, average percentages, and the number of states in each category.
    - **Kamala Strongholds**: 3 states, 4.17M votes.
    - **Trump Strongholds**: 12 states, 10.33M votes.
    - **Swing States**: 36 states, majority of the electoral map, with near-equal average percentages for Kamala (45.5%) and Trump (47.06%).
- **Purpose**: To identify voting trends and the distribution of support across strongholds and competitive regions.

[57]:
```
# Categorize states based on vote percentages
def categorize_state(row):
    if row["Kamala %"] > 60:
        return "Kamala Stronghold"
    elif row["Trump %"] > 60:
        return "Trump Stronghold"
    else:
        return "Swing State"

copy_data_2024["Category"] = copy_data_2024.apply(categorize_state, axis=1)

# Provide summary statistics for each category
category_summary = copy_data_2024.groupby("Category").agg({
    "Kamala Votes": "sum",
    "Trump Votes": "sum",
    "Kamala %": "mean",
    "Trump %": "mean",
    "STATE": "count"
}).rename(columns={"STATE": "Number of States"})

# Save the summary to a CSV
category_summary.to_csv("category_summary_2024_Q2.csv")
print("\nCategory Summary Statistics (2024 Elections):")
print(category_summary)

data_2024.head()
copy_data_2024.head()
```

```
Category Summary Statistics (2024 Elections):
                  Kamala Votes  Trump Votes   Kamala %    Trump %  \
Category
Kamala Stronghold     4170909.0    2568970.0  62.466667  34.433333
Swing State          64244101.0   64046111.0  45.525000  47.063889
Trump Stronghold      5442685.0   10328463.0  32.691667  65.333333

                  Number of States
Category
Kamala Stronghold                3
Swing State                     36
Trump Stronghold                12
```

[57]:

	STATE	Kamala Votes	Kamala %	Kamala EV	Trump Votes	Trump %	Trump EV	State_Code	Category
0	Alabama	769391.0	34.1	0	1457704.0	64.6	9.0	AL	Trump Stronghold
1	Alaska	139812.0	41.4	0	184204.0	54.5	3.0	AK	Swing State
2	Arizona	1582860.0	46.7	0	1770242.0	52.2	11.0	AZ	Swing State
3	Arkansas	396905.0	33.5	0	759241.0	64.2	6.0	AR	Trump Stronghold

	Category	Kamala Votes	Trump Votes	Kamala %	Trump %	Number of States
1	Kamala Stronghold	4170909.0	2568970.0	62.46666666666667	34.43333333333333	3
2	Swing State	64244101.0	64046111.0	45.525000000000006	47.06388888888889	36
3	Trump Stronghold	5442685.0	10328463.0	32.69166666666667	65.33333333333333	12

3. Comparing the 2020 and 2024 Elections

- **Unit of Analysis**: Percentage changes in votes, vote margin changes, and shifts in Electoral Votes between 2020 and 2024.
- **Computation**: The datasets were merged to calculate:
    - **% Change**: Kamala-Biden % Change and Trump % Change by state.
    - **Vote Change**: Kamala-Biden Vote Change and Trump Vote Change by state.
    - **Key Observations**:
        - Kamala Harris experienced notable declines in states like California.
        - Trump gained ground in critical swing states like Arizona and Florida.
    - The comparison highlighted a Democratic loss in EVs (345 for Biden in 2020 vs. 266 for Kamala in 2024) and a Republican gain in both votes and percentages.
- **Purpose**: To identify key shifts in voter behavior and demographic support for both parties across two election cycles.

```python
# Merge 2020 and 2024 datasets on State_Code
merged_data = pd.merge(data_2020, data_2024, on="State_Code", how="outer", suffixes=("_2020", "_2024"))

# Calculate percentage and vote changes
merged_data["Kamala-Biden % Change"] = merged_data["Kamala %"] - merged_data["Biden %"]
merged_data["Trump % Change"] = merged_data["Trump %_2024"] - merged_data["Trump %_2020"]

merged_data["Kamala-Biden Vote Change"] = merged_data["Kamala Votes"] - merged_data["Biden Votes"]
merged_data["Trump Vote Change"] = merged_data["Trump Votes_2024"] - merged_data["Trump Votes_2020"]

# Drop the last row using iloc
merged_data = merged_data.iloc[:-1]


# Save the merged data with changes to a CSV
merged_data.to_csv("merged_2020_2024_elections.csv", index=False)
print("\nMerged 2020 and 2024 Election Data with Changes:")
print(merged_data.head())
```

```
Merged 2020 and 2024 Election Data with Changes:
   STATE_2020  Biden Votes  Biden %  Biden EV  Trump Votes_2020  Trump %_2020  \
0     Alabama     849624.0    36.57       0.0         1441170.0         62.03
1      Alaska     153778.0    42.77       0.0          189951.0         52.83
2     Arizona    1672143.0    49.36      11.0         1661686.0         49.06
3    Arkansas     423932.0    34.78       0.0          760647.0         62.40
4  California   11110250.0    63.48      55.0         6006429.0         34.32

   Trump EV_2020 State_Code  STATE_2024  Kamala Votes  Kamala %  Kamala EV  \
0            9.0         AL     Alabama      769391.0      34.1        0.0
1            3.0         AK      Alaska      139812.0      41.4        0.0
2            0.0         AZ     Arizona     1582860.0      46.7        0.0
3            6.0         AR    Arkansas      396905.0      33.5        0.0
4            0.0         CA  California     9183800.0      58.6       54.0

   Trump Votes_2024  Trump %_2024  Trump EV_2024  Kamala-Biden % Change  \
0         1457704.0          64.6            9.0                  -2.47
1          184204.0          54.5            3.0                  -1.37
2         1770242.0          52.2           11.0                  -2.66
3          759241.0          64.2            6.0                  -1.28
4         5988823.0          38.2            0.0                  -4.88

   Trump % Change  Kamala-Biden Vote Change  Trump Vote Change
0            2.57                  -80233.0            16534.0
1            1.67                  -13966.0            -5747.0
2            3.14                  -89283.0           108556.0
3            1.80                  -27027.0            -1406.0
4            3.88                -1926450.0           -17606.0
```

	STATE_2020	Biden Votes	Biden %	Biden EV	Trump Votes_2020	Trump %_2020	Trump EV_2020	State_Code	STATE_2024	K
1	Alabama	849624.0	36.57	0.0	1441170.0	62.03	9.0	AL	Alabama	
2	Alaska	153778.0	42.77	0.0	189951.0	52.83	3.0	AK	Alaska	
3	Arizona	1672143.0	49.36	11.0	1661686.0	49.06	0.0	AZ	Arizona	
4	Arkansas	423932.0	34.78	0.0	760647.0	62.4	6.0	AR	Arkansas	
5	California	11110250.0	63.48	55.0	6006429.0	34.32	0.0	CA	California	

Kamala %	Kamala EV	Trump Votes_2024	Trump %_2024	Trump EV_2024	Kamala-Biden % Cha…	Trump % Change	Kamala-Biden Vote C…	Trump Vote Change
34.1	0.0	1457704.0	64.6	9.0	-2.469999999999999	2.569999999999993	-80233.0	16534.0
41.4	0.0	184204.0	54.5	3.0	-1.3700000000000045	1.6700000000000017	-13966.0	-5747.0
46.7	0.0	1770242.0	52.2	11.0	-2.6599999999999966	3.1400000000000006	-89283.0	108556.0
33.5	0.0	759241.0	64.2	6.0	-1.2800000000000011	1.8000000000000043	-27027.0	-1406.0
58.6	54.0	5988823.0	38.2	0.0	-4.8799999999999955	3.8800000000000026	-1926450.0	-17606.0

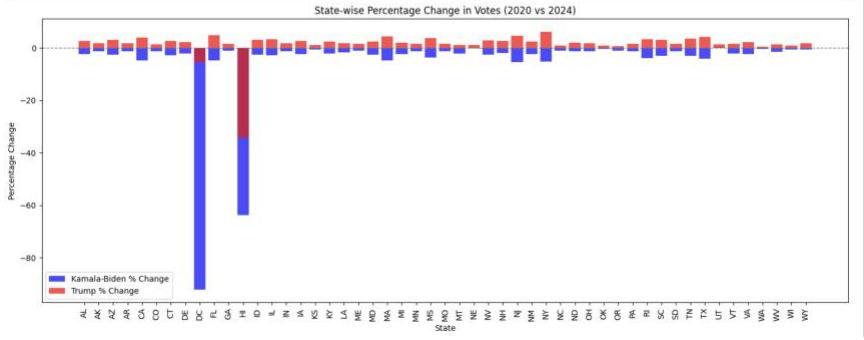
Other Ways of Analysis:

1. Percentage Change in Voter Support (2020 vs. 2024)

- **Graph**: State-wise percentage change for Kamala Harris and Donald Trump.
- **Findings**:
  - Significant declines for Kamala Harris in states like D.C. and California.
  - Gains for Trump in states like Florida and Texas, indicating a shift toward Republican support.

```python
import matplotlib.pyplot as plt

# Plot percentage change for Kamala and Trump
plt.figure(figsize=(15, 6))
plt.bar(merged_data["State_Code"], merged_data["Kamala-Biden % Change"], label="Kamala-Biden % Change", alpha=0.7, color='blue')
plt.bar(merged_data["State_Code"], merged_data["Trump % Change"], label="Trump % Change", alpha=0.7, color='red')
plt.axhline(0, color='gray', linestyle='--', linewidth=1)
plt.title("State-wise Percentage Change in Votes (2020 vs 2024)")
plt.xlabel("State")
plt.ylabel("Percentage Change")
plt.xticks(rotation=90)
plt.legend()
plt.tight_layout()
plt.show()
```



2.Electoral Map Visualization (2024 Presidential Results)

- **Map Analysis**:
  - Kamala Harris dominated in the west and northeast.
  - Trump secured strongholds in southern and midwestern states.
- **Key Insights**:
  - Swing states remain decisive.
  - State-wise performance helps in strategizing future campaigns.

```python
# Determine the winner for each state
def determine_winner(row):
    if row["Kamala %"] > row["Trump %_2024"]:
        return "Biden/Harris"
    else:
        return "Trump"

merged_data["Winner"] = merged_data.apply(determine_winner, axis=1)

# Create the map using Plotly Express
fig = px.choropleth(
    merged_data,
    locations="State_Code",  # Column containing state abbreviations
    locationmode="USA-states",  # Map to USA states
    color="Winner",  # Column containing the winner (Biden/Harris or Trump)
    hover_name="STATE_2020",  # Info to show on hover
    scope="usa",  # Map scope limited to the USA
    color_discrete_map={"Trump": "red", "Biden/Harris": "blue"},  # Assign colors
)

# Add state labels on the map
for i, row in merged_data.iterrows():
    fig.add_scattergeo(
        locationmode='USA-states',
        locations=[row["State_Code"]],
        text=row["State_Code"],
        mode='text',
        showlegend=False
    )

# Update layout for better visualization
fig.update_layout(
    title_text="2024 U.S. Presidential Election Results",
    geo=dict(showlakes=True, lakecolor="lightblue"),
)

# Show the map
fig.show()
```

Description of the Program:

The analysis was conducted using a Jupyter Notebook file named **presidential_elections_analysis.ipynb**. This notebook systematically explores and processes two datasets representing the 2020 and 2024 U.S. presidential elections. The program performs the following tasks:

1. **Data Loading and Cleaning**:
   - Loads the 2020 and 2024 datasets, fills missing values (e.g., Electoral Votes), and standardizes column names.
   - Merges the two datasets on the State_Code column to enable comparison between election years.
2. **Analysis**:
   - **Q1**: Summarizes the 2024 election data, including total votes, average percentages, and Electoral Votes (EVs) for both candidates.
   - **Q2**: Categorizes the 2024 data into Kamala Strongholds, Trump Strongholds, and Swing States and calculates summary statistics for each category.
   - **Q3**: Computes percentage and vote changes between 2020 and 2024 for each state, highlighting shifts in voter preferences.
   - **Q4**: Compares the aggregated performance of Democrats and Republicans between the two elections and visualizes changes using maps and graphs.
3. **Output Generation**:
   - Creates structured output files for each analysis question.
4. **Visualization**:
   - Includes bar plots for percentage changes and an interactive U.S. map to represent state-wise election results for 2024.

Description of the Output Files:

The program generates the following four output files, each corresponding to one of the main analysis questions:

1. **Summary_df_2024_Q1.csv**:
   - Contains the summarized data for the 2024 elections, including total votes, average percentages, and Electoral Votes for Kamala Harris and Donald Trump.
2. **category_summary_2024_Q2.csv**:
   - Provides category-wise summary statistics for 2024 elections:

        - Total votes, average percentages, and the number of states in each category (Kamala Strongholds, Trump Strongholds, and Swing States).
3. **merged_2020_2024_elections.csv:**
    - Contains the merged dataset for 2020 and 2024 elections, with additional columns:
        - Percentage and vote changes for Kamala-Biden and Trump between the two elections.
4. **comparison_2020_2024_summary.csv:**
    - Includes a comparative analysis of the 2020 and 2024 elections, showing aggregate performance (votes, percentages, EVs) for each candidate in both years.