**EDA ELECTIONS**

**1. Dataset Overview**

* **Total Records**: The dataset consists of **1,600** entries (respondents).
* **Columns**: There are **16 columns** representing various attributes such as demographics, voting intentions, media consumption, and more.

**2. Age Distribution**

* **Range**: The age of respondents ranges from **18** to **90** years.
* **Average Age**: The average age is approximately **50** years.
* **Distribution**: The distribution is slightly skewed towards older individuals, with a noticeable concentration in the **50-60** age range.

**3. Education Levels**

* **Categories**: The dataset includes various education levels, from **incomplete school education** to **advanced degrees**.
* **Distribution**: The data shows a diverse range of educational backgrounds, with a higher proportion of respondents having **secondary** and **higher education**.

**4. Missing Data Analysis**

* **Job Type**: The job\_type column has **908** missing values, which is **56.75%** of the total data.
* **Company Type**: The company\_type column has **721** missing values, representing **45.06%** of the data.
* **Impact**: The significant missing data in these columns suggests that any analysis involving these variables must be handled carefully, potentially through imputation or exclusion.

**5. Correlation Analysis**

* **Overall Correlations**: The correlation analysis between numerical variables, such as age and weight1, did not reveal any extremely strong correlations.
* **Key Observation**: This suggests that the numerical variables in the dataset provide unique information without much overlap.

**6. Federal District and Candidate Preference**

* **Variation Across Districts**: Candidate preferences vary across different federal districts, indicating regional differences in political support.
* **Potential Insight**: This variation highlights the importance of geographical factors in understanding voting patterns.
* Putin is the most favoured candidate across all the federal districts.

**7. Media Usage**

* **Television vs. Internet**: The dataset provides insight into how respondents use television and the internet.
* **Frequency**: There’s a noticeable variation in the frequency of television and internet usage, which could correlate with factors like age, education, and income.
* Most people watch television and surf the internet for less than 4 hours everyday respectively.
* **Comparison**: The analysis suggests that certain demographic groups might prefer one medium over the other, but further analysis would be needed to confirm this.

**8. Bivariate Analysis (Age vs. Income)**

* **Income Distribution**: The bivariate analysis of age and income suggests that income distribution varies across different age groups.
* **Key Trend**: There might be a trend where certain age groups tend to fall into specific income brackets, with younger respondents generally falling into lower income brackets, and middle-aged respondents in higher brackets.

**9. Crosstab and Heatmap Analysis**

* **Education and Candidate Preference**: A crosstab between education and candidate preference was created, showing how educational background relates to candidate choice.
* **Most of the candidates from all the educational levels respectively chose Putin**.
* **Key Finding**: The heatmap visualization of this crosstab highlights significant patterns, such as candidates being more popular among certain education levels.