

Carrer Aspiration Responses

Quick Analysis of the dataset :

```
[7]: df.isna().sum()
```

[7]:	Timestamp	0
	Your Current Country.	0
	Your Current Zip Code / Pin Code	0
	Your Gender	0
	Which of the below factors influence the most about your career aspirations ?	0
	Would you definitely pursue a Higher Education / Post Graduation outside of India ? If only you have to self sponsor it.	0
	How likely is that you will work for one employer for 3 years or more ?	0
	Would you work for a company whose mission is not clearly defined and publicly posted.	0
	How likely would you work for a company whose mission is misaligned with their public actions or even their product ?	0
	How likely would you work for a company whose mission is not bringing social impact ?	0
	What is the most preferred working environment for you.	0
	Which of the below Employers would you work with.	0
	Which type of learning environment that you are most likely to work in ?	0
	Which of the below careers looks close to your Aspirational job ?	0
	What type of Manager would you work for without looking into your watch ?	0
	Which of the following setup you would like to work ?	0
	Would you work for a company that has Laid-Off Employees recently ?	381
	How likely is that you will work for one employer for 7 years or more ?	387
	Email address	1849
	For the First 3 years of your career, what is your minimum expected monthly salary in Hand?	0
	After 5 years of your career from now, what is your minimum expected monthly salary in Hand?	0
	Would you work for a company that has NO-REMOTE working policy	3441
	What is your minimum expectations of salary when you are starting your career (Monthly in hand salary)	3441
	What Kind of company would you like to work	3441
	Would you work under a manager who uses abusive words at workplace or if you get to know the manager has done that in the past	3441
	How many hours would you be willing to work every day	3441
	How often do you need a full week break, to stay healthy with the work life balance	3441
	What would make you happier and productive at work	3441
	What would frustrate you at work ?	3441
	dtype: int64	

[8]:	Timestamp	0.00%
	Your Current Country.	0.00%
	Your Current Zip Code / Pin Code	0.00%
	Your Gender	0.00%
	Which of the below factors influence the most about your career aspirations ?	0.00%
	Would you definitely pursue a Higher Education / Post Graduation outside of India ? If only you have to self sponsor it.	0.00%
	How likely is that you will work for one employer for 3 years or more ?	0.00%
	Would you work for a company whose mission is not clearly defined and publicly posted.	0.00%
	How likely would you work for a company whose mission is misaligned with their public actions or even their product ?	0.00%
	How likely would you work for a company whose mission is not bringing social impact ?	0.00%
	What is the most preferred working environment for you.	0.00%
	Which of the below Employers would you work with.	0.00%
	Which type of learning environment that you are most likely to work in ?	0.00%
	Which of the below careers looks close to your Aspirational job ?	0.00%
	What type of Manager would you work for without looking into your watch ?	0.00%
	Which of the following setup you would like to work ?	0.00%
	Would you work for a company that has Laid-Off Employees recently ?	9.71%
	How likely is that you will work for one employer for 7 years or more ?	9.86%
	Email address	47.13%
	For the First 3 years of your career, what is your minimum expected monthly salary in Hand?	0.00%
	After 5 years of your career from now, what is your minimum expected monthly salary in Hand?	0.00%
	Would you work for a company that has NO-REMOTE working policy	87.71%
	What is your minimum expectations of salary when you are starting your career (Monthly in hand salary)	87.71%
	What Kind of company would you like to work	87.71%
	Would you work under a manager who uses abusive words at workplace or if you get to know the manager has done that in the past	87.71%
	How many hours would you be willing to work every day	87.71%
	How often do you need a full week break, to stay healthy with the work life balance	87.71%
	What would make you happier and productive at work	87.71%
	What would frustrate you at work ?	87.71%
	dtype: object	

```
[47]: df.columns
```

```
[47]: Index(['Timestamp', 'Current_Country', 'Current_Zip_Code', 'Gender',  
        'Career_Influence', 'Higher_Education_Outside_India',  
        'Employer_3_Years_Likelihood', 'Undefined_Mission_Company',  
        'Misaligned_Mission_Company', 'No_Social_Impact_Company',  
        'Preferred_Working_Environment', 'Preferred_Employers',  
        'Learning_Environment', 'Aspirational_Career', 'Preferred_Manager',  
        'Work_Setup', 'Laid_Off_Company', 'Employer_7_Years_Likelihood',  
        'Email_address', 'Career_Initial_Salary', 'Career_5_Years_Salary',  
        'No_Remote_Working_Policy', 'Starting_Salary_Expectation',  
        'Preferred_Company_Type', 'Abusive_Manager', 'Hours_Per_Day',  
        'Full_Week_Break_Frequency', 'Happiness_At_Work',  
        'Frustration_At_Work'],  
        dtype='object')
```

```
[48]: len(df.columns)
```

In my project, I choose not to drop any rows with missing data to maintain the integrity of the dataset, especially since the project description required handling missing values without removal. Here's how I handled the missing values:

Step-by-Step Explanation of Data Cleaning and Handling Missing Data:

1. Problem: Missing Data

- ✓ **Issue:** The dataset contained several missing values across various columns. This is a common issue in real-world data, and addressing it properly is critical for ensuring the accuracy and quality of the analysis or machine learning models.
- ✓ **Solution:** I first identified which columns had missing values using `df.isna().sum()`. Then, I calculated the percentage of missing values for each column to prioritize which ones needed to be handled first.

2. Handling Missing Categorical Columns:

- ✓ **Problem:** Many categorical columns, such as "Would you work for a company that has Laid-Off Employees recently?" and "How likely is that you will work for one employer for 7 years or more?", had missing values that could affect analysis.
- ✓ **Solution:** I replaced these missing values with placeholders like 'Unknown' or 'No Response'. This ensured the data remained usable while still clearly indicating where information was missing.

3. Handling Missing Numeric Columns:

- ✓ **Problem:** Some numeric columns, such as salary expectations or work hours, had missing values that could disrupt analysis.
- ✓ **Solution:** I replaced missing values in these columns with sensible defaults:
 - **Salary columns:** Filled with 0 for missing salary data.
 - **Work hours:** I replaced missing values with 0, then handled the column's format by extracting numeric values (in case it had text) using `str.extract()`.

4. Special Case: Email Address Column:

- ✓ **Problem:** The "Email address" column had a significant number of missing values.
- ✓ **Solution:** I replaced missing email entries with 'No Email', as it was a meaningful placeholder to indicate the absence of an email address without affecting the analysis.

5. Handling Other Categorical and Text Columns:

- ✓ **Problem:** Columns like "What Kind of company would you like to work" and "What would frustrate you at work?" had missing values that needed consistent treatment.
- ✓ **Solution:** I replaced missing values in these columns with appropriate placeholders like 'No Preference' or 'Not Specified'. This ensures that these fields are still useful for analysis and keeps the dataset consistent.

6. Datetime Conversion:

- ✓ **Problem:** The "Timestamp" column was in a non-standard format, which could interfere with any time-based analysis or feature engineering.
- ✓ **Solution:** I converted this column into a proper datetime format using `pd.to_datetime()` to ensure it was ready for time-based calculations.

7. Renaming Columns for Clarity:

- ✓ **Problem:** Column names were unclear or had extra characters.
- ✓ **Solution:** Renamed columns to make them more readable and descriptive using `df.rename()`.

8. Cleaning Data Values:

- ✓ Problem: The data contained inconsistent or non-standard terms.
- ✓ Solution: Mapped values in specific columns using predefined dictionaries to standardize the values (e.g., career influences, higher education preferences).

9. Handling Work Setup Data:

- ✓ Problem: Work setup data was mixed and inconsistent.
- ✓ Solution: Created a custom function to categorize the work setup values and apply it to the column.

10. Removing Unwanted Characters:

- ✓ Problem: Some rows contained unwanted strings (e.g., "Xxxxx", "Xxxx", "-").
- ✓ Solution: Identified these values and replaced them with meaningful values or the mode of the column.

11. Handling Empty or Blank Values:

- ✓ Problem: Columns contained blank values or invisible characters.
- ✓ Solution: Used `str.strip()` to remove leading/trailing spaces or tabs, then replaced empty values with the mode (most common value) in the `Current_Zip_Code` column.

12. Final Check and Data Integrity:

- ✓ **Problem:** After cleaning the data, I needed to verify that all the transformations were applied correctly and that no data had been lost in the process.
- ✓ **Solution:** I checked the unique values in columns after handling missing values to ensure the transformations were accurate. I also made sure that no critical information was omitted, and that the dataset was now clean and ready for analysis.