Data

**Data Source & method of collection:**

I chose to use a publicly available dataset from Kaggle called: “[Data Science Salaries 2023](https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023?resource=download)”

**This dataset has been sourced from aijobs.net, however the method of collection is not disclosed.**

**Data Contents:**

**This dataset contains the Salaries of Different Data Science Fields in the Data Science Domain.**

**Data Profile:**

**A few minor formatting adjustments were made to make the data clearer.**

**1171 duplicate values were found and removed in Excel.**

**After duplicates were removed, the dataset is left with 2585 rows and 11 columns.**

**Limitations & Ethics:**

**The dataset does not have any significant limitations. It also does not contain any PII that could pose ethical risks.**

**Possible Hypotheses:**

* **Null Hypothesis: There is no significant difference in salaries across different experience levels.**

**Alternative Hypothesis: Salaries vary significantly across different experience levels.**

* **Null Hypothesis: There is no significant relationship between remote work ratio and salary.**

**Alternative Hypothesis: There is a significant relationship between remote work ratio and salary.**

* **Null Hypothesis: There is no significant difference in salaries based on employee residence.**

**Alternative Hypothesis: Salaries vary significantly based on employee residence.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Index** | **Columns** | **Description** | **Time Variant/**  **Invariant** | **Data Type** |
| **1** | **work\_year** | **The year the salary was paid.** | **Variant** | **Quantitative** |
| **2** | **experience\_level** | **The experience level in the job during the year.** | **Variant** | **Qualitative** |
| **3** | **employment\_type** | **The type of employment for the role.** | **Invariant** | **Qualitative** |
| **4** | **job\_title** | **The role worked in during the year.** | **Variant** | **Qualitative** |
| **5** | **salary** | **The total gross salary amount paid.** | **Variant** | **Quantitative** |
| **6** | **salary\_currency** | **The currency of the salary paid as an ISO 4217 currency code.** | **Invariant** | **Qualitative** |
| **7** | **salary\_in\_usd** | **The salary in USD (converted amount).** | **Variant** | **Quantitative** |
| **8** | **employee\_residence** | **Employee's primary country of residence during the work year as an ISO 3166 country code.** | **Variant** | **Qualitative** |
| **9** | **remote\_ratio** | **The overall amount of work done remotely.** | **Variant** | **Quantitative** |
| **10** | **company\_location** | **The country of the employer's main office or contracting branch.** | **Variant** | **Qualitative** |
| **11** | **company\_size** | **The median number of people that worked for the company during the year.** | **Invariant** | **Quantitative** |