

# Data Cleaning Documentation

## Overview

This document details the step-by-step cleaning process applied to the Profession Survey Dataset to ensure accuracy, consistency, and usability for analysis.

## Data Cleaning Steps

### 1. Removed Blank Rows

- Identified and deleted 284 blank rows to maintain data integrity.

### 2. Standardized Job Titles

- Ensured uniformity by correcting inconsistent job titles.
- Removed duplicate and irrelevant job titles.

### 3. Categorized Job Titles

- Mapped job titles into well-defined categories to streamline analysis.
- Ensured all data-related roles were classified correctly.

### 4. Standardized Industry Names

- Merged similar industry names (e.g., Finance and Insurance were kept separate, while Staffing & Recruiting were merged).
- Moved miscellaneous industries into appropriate existing categories instead of "Other."

### 5. Handled Salary Ranges

- Converted salary ranges from text format (e.g., "106k-125k") to numerical values using the SUBSTITUTE function to remove 'k'.
- Split salary ranges into minimum and maximum values using TEXT TO COLUMNS.
- Calculated the average salary per row using the formula:  
`=AVERAGE(MIN_VALUE, MAX_VALUE)`
- Rounded salary values to the nearest whole number using `ROUND()`.

### 6. Addressed "Unknown" and "Other" Categories

- Re-examined job titles and industries under "Unknown" and "Other."

- Properly assigned valid job titles and industries where possible.
- Reduced reliance on the "Other" category to enhance data clarity.

#### 7. Ranked Industries by Average Salary

- Used pivot tables to filter the Top 10 industries and rank them accordingly.

#### 8. Error Handling

- Resolved #DIV/0! errors in calculations by ensuring no division by zero using:  
*=IFERROR(Calculation, 0)*
- Verified all numeric conversions to prevent inconsistencies.

#### 9. Final Review and Data Validation

- Conducted a final review to ensure all cleaning steps were correctly applied.
- Verified that every industry and job title had meaningful assignments.
- Checked consistency in salary values and rankings.

---

This cleaned dataset is now structured and ready for in-depth analysis and visualization.

---