Exploratory Data Analysis (EDA)

Visualizing and Interpreting Job Market Trends

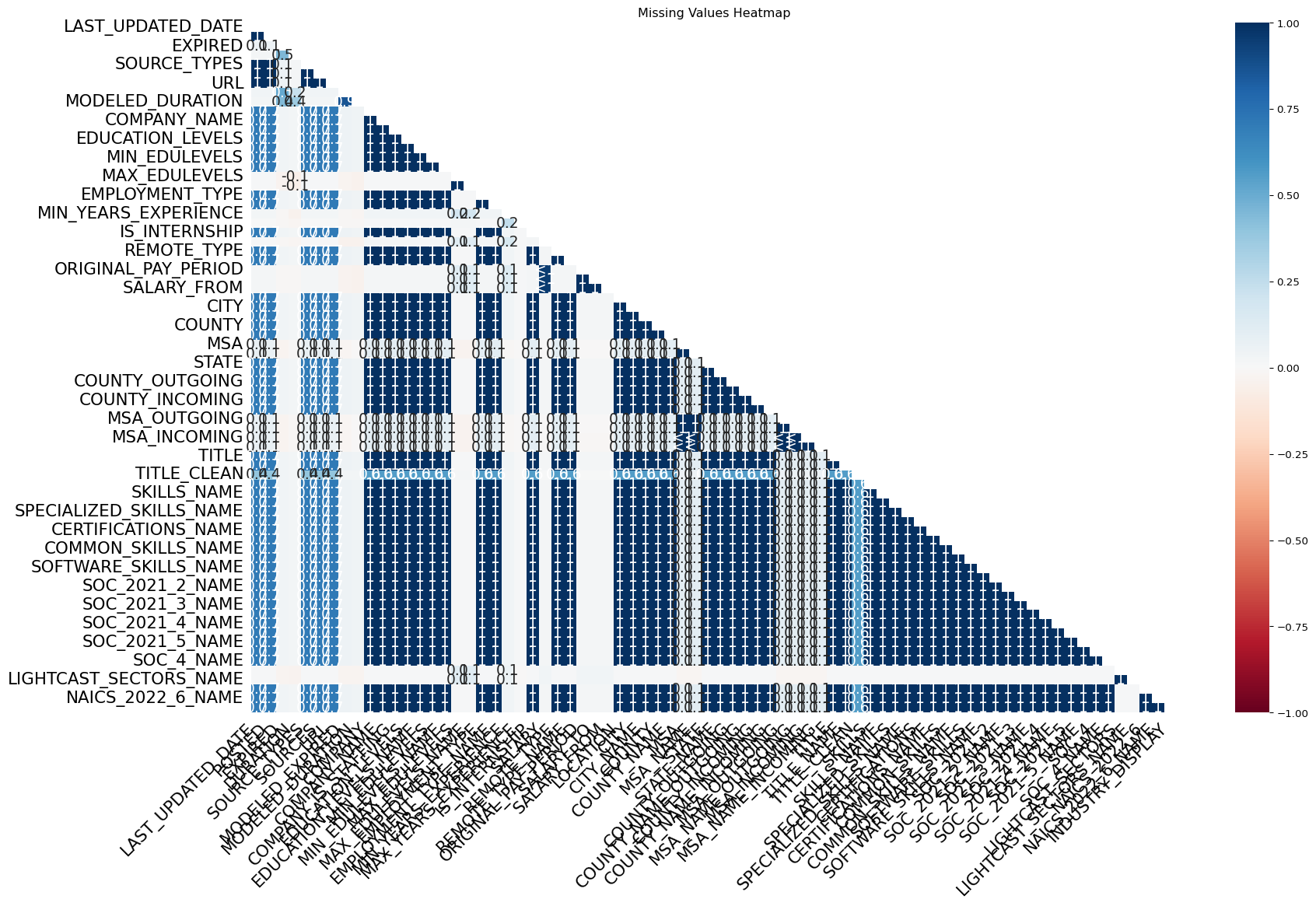
Connor Coulter

Wei Wang

Balqis Bevi Abdul Hannan Kanaga

Loaded dataset: (72498, 131)  
Derived non-null: {'INDUSTRY\_DISPLAY': np.int64(72454), 'SALARY\_DISPLAY': np.int64(72498)}  
Remaining columns (first 30): ['LAST\_UPDATED\_DATE', 'POSTED', 'EXPIRED', 'DURATION', 'SOURCE\_TYPES', 'SOURCES', 'URL', 'MODELED\_EXPIRED', 'MODELED\_DURATION', 'COMPANY', 'COMPANY\_NAME', 'COMPANY\_IS\_STAFFING', 'EDUCATION\_LEVELS', 'EDUCATION\_LEVELS\_NAME', 'MIN\_EDULEVELS', 'MIN\_EDULEVELS\_NAME', 'MAX\_EDULEVELS', 'MAX\_EDULEVELS\_NAME', 'EMPLOYMENT\_TYPE', 'EMPLOYMENT\_TYPE\_NAME', 'MIN\_YEARS\_EXPERIENCE', 'MAX\_YEARS\_EXPERIENCE', 'IS\_INTERNSHIP', 'SALARY', 'REMOTE\_TYPE', 'REMOTE\_TYPE\_NAME', 'ORIGINAL\_PAY\_PERIOD', 'SALARY\_TO', 'SALARY\_FROM', 'LOCATION']

/tmp/ipykernel\_3669/3965281858.py:37: FutureWarning:  
  
A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.  
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.  
  
For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

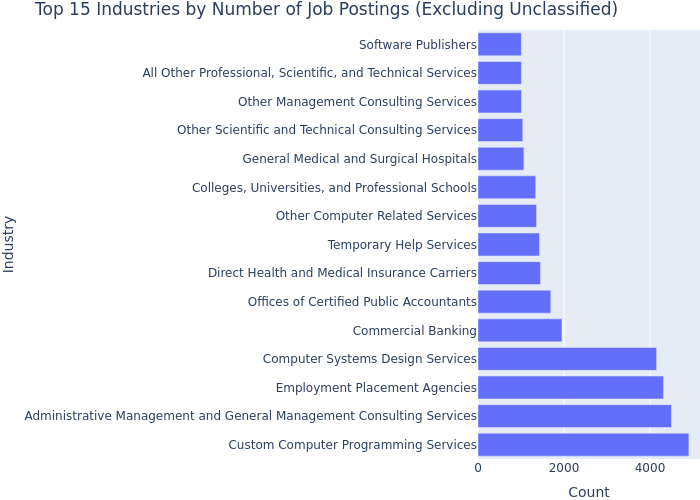


/tmp/ipykernel\_3669/3965281858.py:75: FutureWarning:  
  
A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.  
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.  
  
For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.  
  
  
  
/tmp/ipykernel\_3669/3965281858.py:78: FutureWarning:  
  
A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.  
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.  
  
For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

Removed 3300 duplicates using ['TITLE', 'COMPANY\_NAME', 'LOCATION', 'POSTED']  
Final SALARY\_DISPLAY Non-Null Count: 69198

# Job Postings by Industry (Top 15)

Unable to display output for mime type(s): text/html



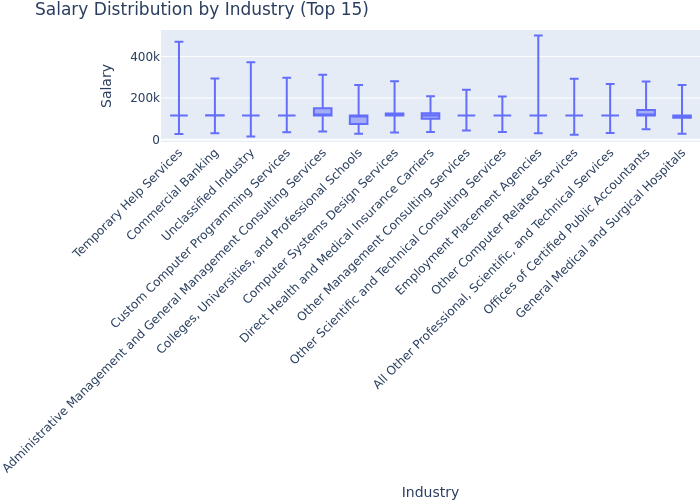
## Rationale

Highlights sectors where demand is concentrated, showing which industries are actively hiring.

## Key Insights

* Top Hiring Industries: Custom Computer Programming, Management Consulting, and Employment Agencies dominate job postings.
* Skewed Distribution: The top 4 industries account for a significantly larger share of job postings than the rest.
* Professional Services Focus: Many high-posting sectors are centered around tech, consulting, healthcare and education — reflecting demand for knowledge-based roles.

# Salary Distribution by Industry (Top 15)



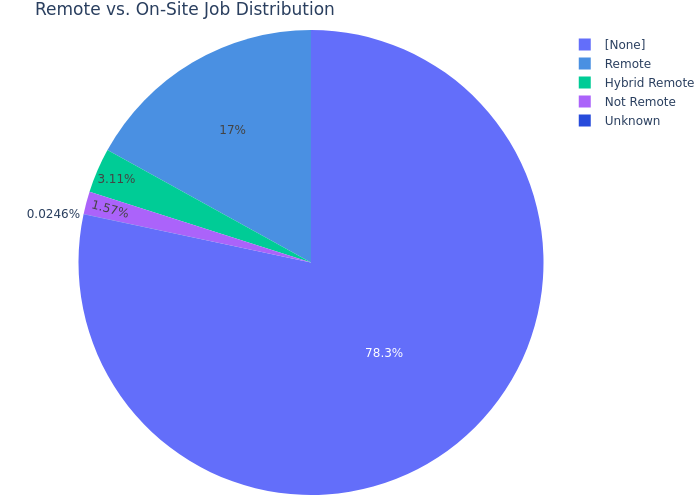
## Rationale

Shows where negotiation power exists and highlights industries paying well.

## Key Insights

* Wide Salary Ranges in Staffing & Tech Services: Industries like Temporary Help Services and Employment Placement Agencies exhibit large salary spreads with high outliers, though their median pay remains modest.
* Stable Pay in Professional Sectors: Most industries maintain a consistent median salary around $100K–$150K, reflecting standardized compensation and less variation in negotiation power.

# Remote vs. On-Site Jobs



## Rationale

Workplace flexibility is a major factor in today’s job market.

## Key Insights

* Limited Remote Availability: Only about 17% of job postings are labeled as Remote, with Hybrid Remote and Not Remote making up even smaller portions.
* Data Gaps in Job Listings: A significant 78.3% of postings lack remote classification, indicating either incomplete employer data or inconsistent labeling, which may affect job seekers’ filtering and selection.