1. Company region and employ residence are regrouped into regions: Africa, Asia, , Americas, Europe, Oceania
2. Over 90 diferent job titles are grouped into 6 groups, keeping the top most frequent and group remaining into ‘others’

|  |  |  |
| --- | --- | --- |
| job\_title | percent | Total |
| Data Engineer | 27.7% | 76.8% |
| Data Scientist | 22.4% |
| Data Analyst | 16.3% |
| Machine Learning Engineer | 7.7% |
| Analytics Engineer | 2.7% |
| others | 23.2% | 23.2% |

|  |  |
| --- | --- |
| Work year | percent |
| 2023 | 47.5% |
| Data Scientist | 44.3% |
| Data Analyst | 6.1% |
| Machine Learning Engineer | 2.0% |

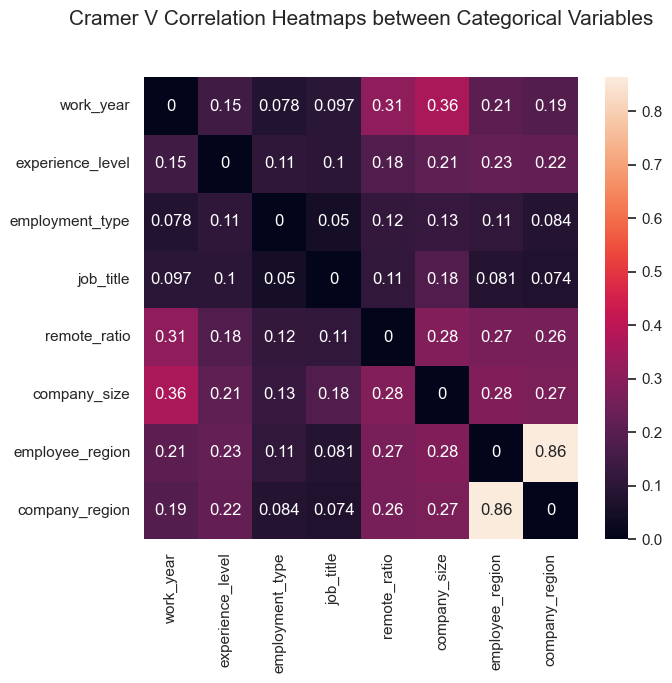
|  |  |
| --- | --- |
| Experience Level | percent |
| Entry Level | 8.5% |
| Medium Level | 21.4% |
| Senior Level | 67.0% |
| Exective Level | 3.0% |

|  |  |
| --- | --- |
| Remote Ratio | percent |
| 0 | 51.2% |
| 50 | 5.0% |
| 100 | 43.8% |

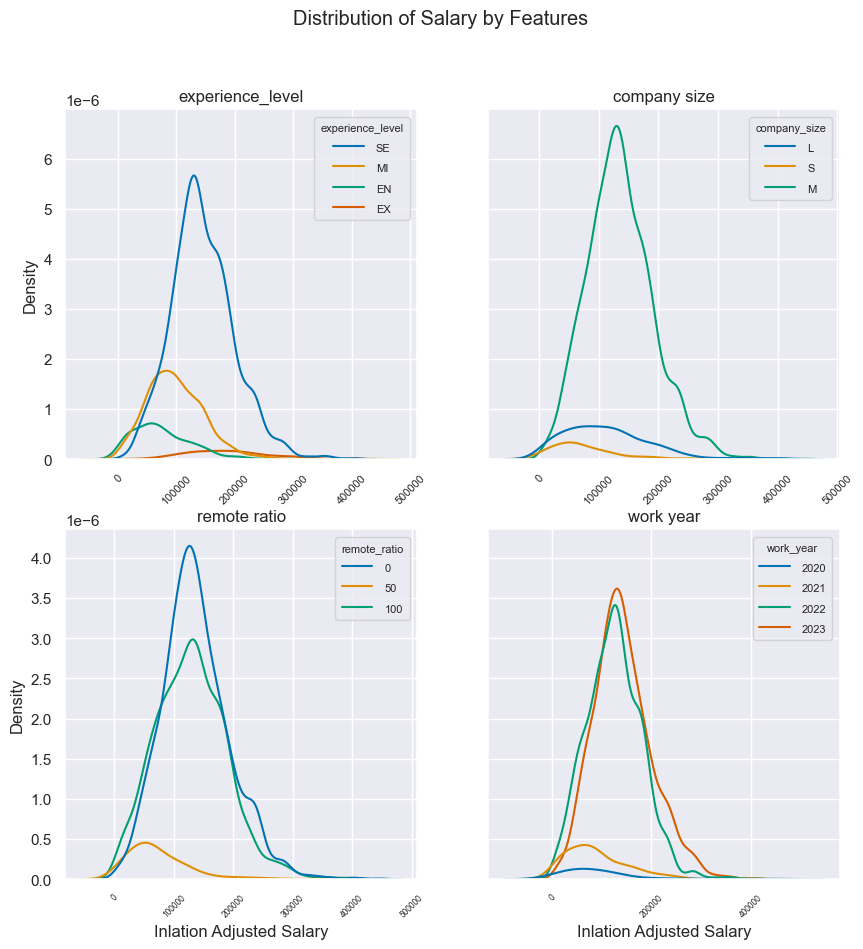
|  |  |
| --- | --- |
| Company Size | percent |
| Small | 3.9% |
| Medium | 84.0% |
| Large | 12.1% |

|  |  |
| --- | --- |
| Employee Region | percent |
| Africa | 0.5% |
| Americas | 83.6% |
| Asia | 3.2% |
| Europe | 12.4% |
| Oceania | 0.4% |

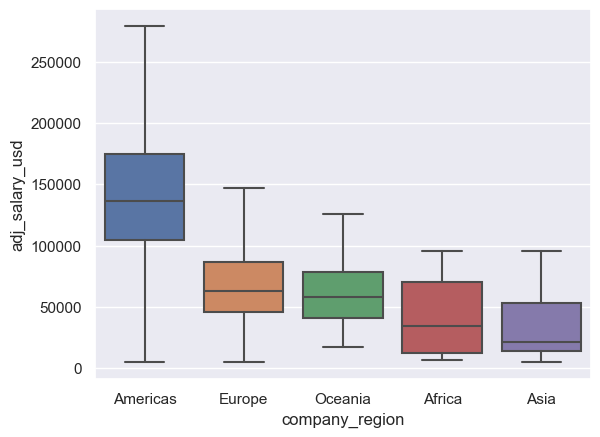
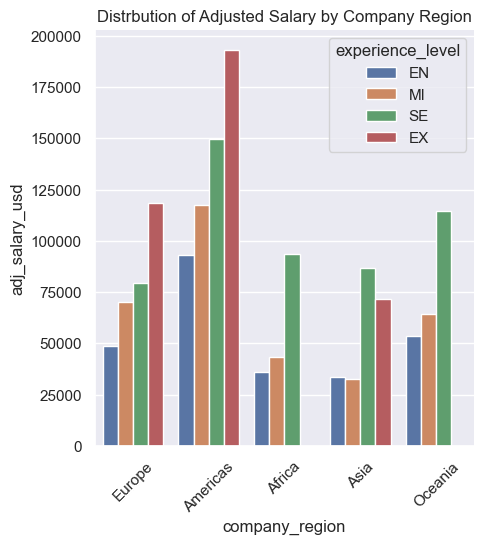
|  |  |
| --- | --- |
| Company Region | percent |
| Africa | 0.4% |
| Americas | 84.4% |
| Asia | 2.6% |
| Europe | 12.2% |
| Oceania | 0.5% |

1. Cramer’s V correlation between Employee\_region and company\_region is 0.86. out of 3755 records, ~98.5% having Employee\_region same as company\_region, and only 58 records Employee\_region and company\_region are not the same. See Cramer’s V heatmap. 
2. Visual exploration of influence of factors on salary.

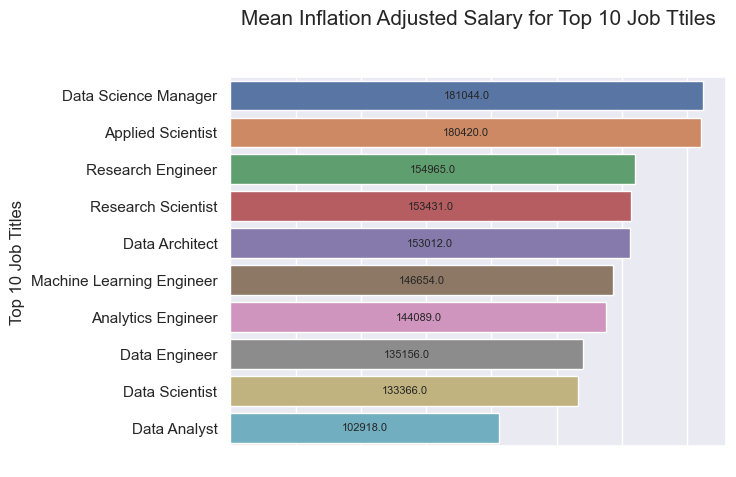
* As exepcted, more experience means higher earning. Executives make the least, which might be because administration does not pay as high.The highest earning are people with senior level experience.
* Medium-sized companies pay significantly more, followed by large-sized and then small companies.
* In the past four years (2020-2023), pay for dada science jobs has increased quite a bit, with the big leap happened going from 2021 to 2022).
* Working exclusively on-site or exclusively remote both get paid more than hybrid, with on-site slightly higher than remote. The reason could be people work hybrid only work part-time and get paid less due to shorter hour



* Compsny region matters too. Among the five regions where our data science job records were collected, Americas, Europe,Oceania, Africa and Asia, salary goes in an decreasing order, with Americas pay employee significantly higher than the rest four. Nonetheless, more experience means higher pay regardless of region.

* Top 10 most common/popular data science jobs are highest paid jobs are Data Engineer,Data Scientist, Data Analyst, Machine Learning Engineer, Analytics Engineer, Data Architect, Research Scientist, Data Science Manager, Applied Scientist, Research Engineer. Their average annual are listed below in a decreasing order.



**Modeling**

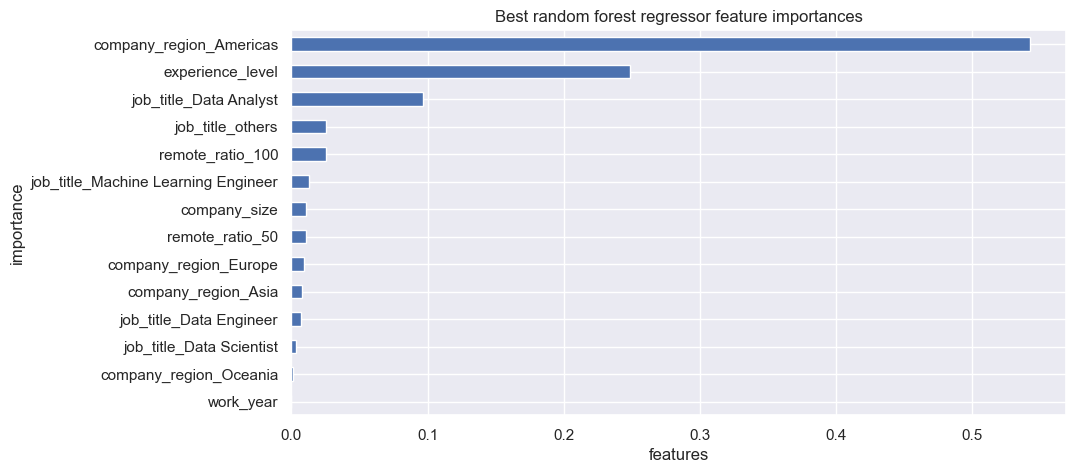
Random Forest regression was applied and hyperparameter are tuned by GridSearchCV. All features were entered as categorical.

GridsearchCV best parameters :

{'max\_depth': 5, 'n\_estimators': 16}

GridsearchCV R-squared:0.371

GridsearchCV Adjusted R-squared:0.362



**Discussion**