Dataset: DataAnalyst.csv https://www.kaggle.com/andrewmvd/data-analyst-jobs

Objective:
Using the dataset, I will explore answers for:

1. What kind of titles get higher salaries?
2. What companies pay more?
3. Does job location matter for salary variations?

Methodologies:

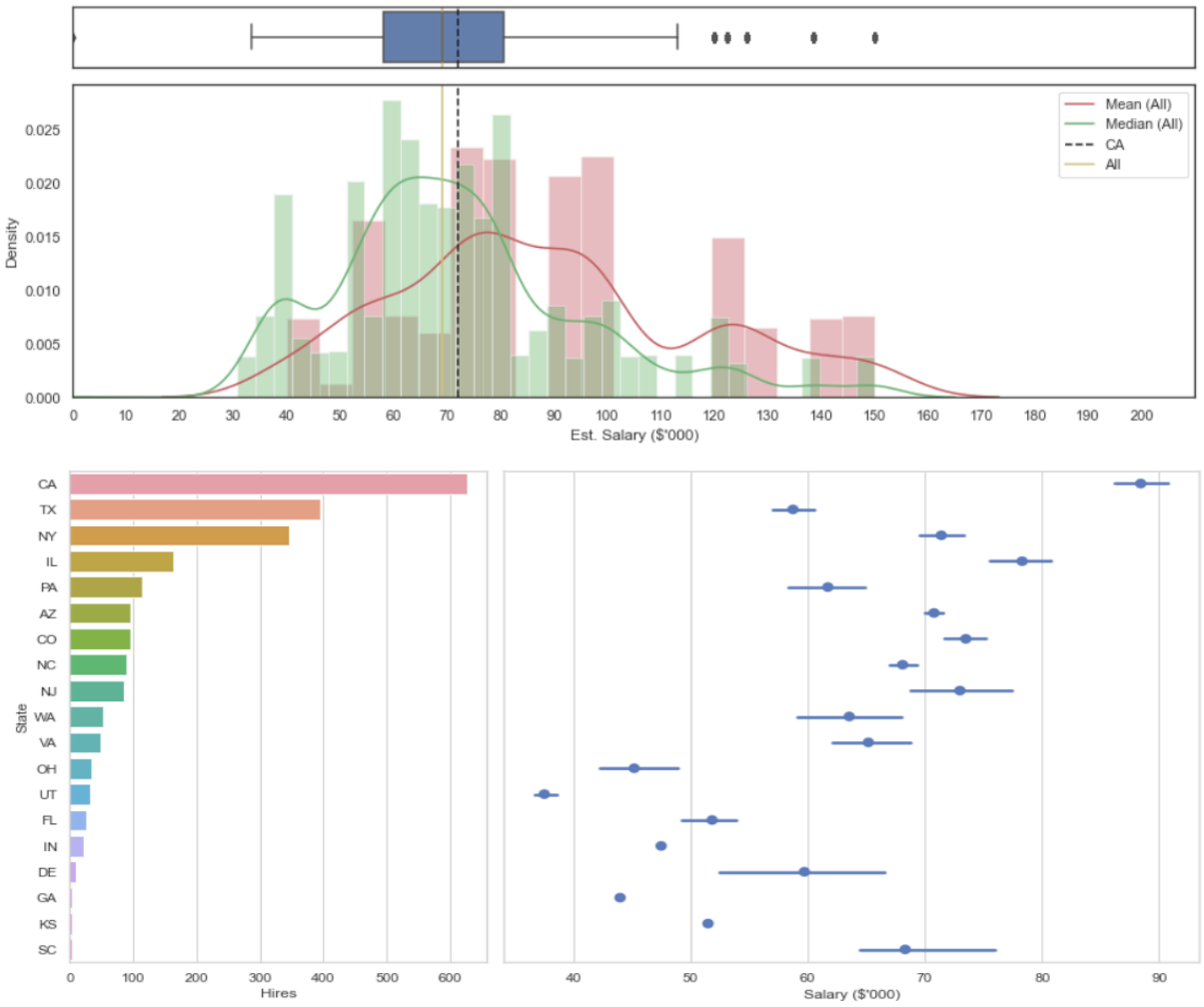
1. Exploratory Data Analysis (distributions, boxplots, scatterplots, heatmaps, etc)
2. T-test
3. Regression

Limitations:

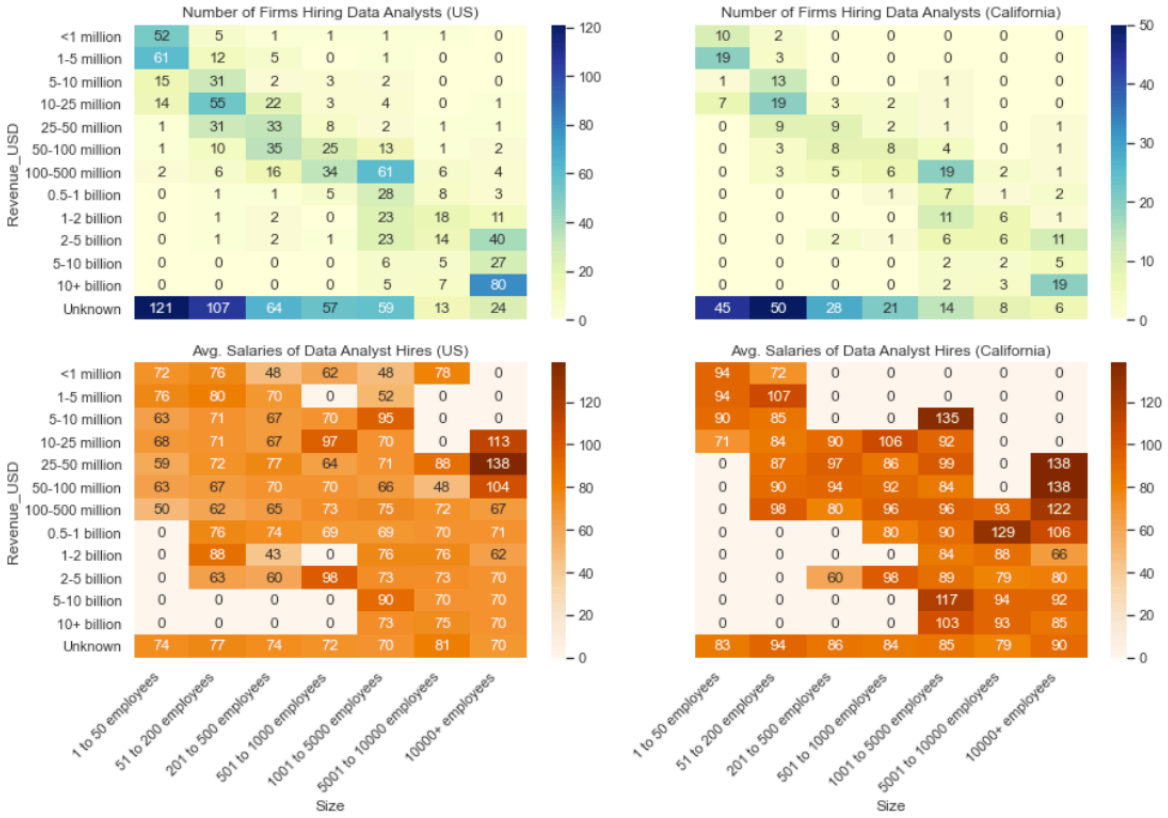
1. Dataset is sourced from Glassdoor, and salaries are estimates per Glassdoor (*data accuracy - a better source could have been payroll processing systems like ADP; source of truth is IRS data*)
2. Data is dated July 2020, and has 27 positions marked as 'remote' – this is 4 months into pandemic (*I am concerned with data quality/capture issues at source may affect my analysis)*
3. Salary (estimates) are nominal and not adjusted by living costs and price indices
4. Dataset is assumed to be representative of actual job market traits
5. The dataset is mostly categorical data, I will skip PMF and CDF analysis.

Key Observations:

1. Job location is the main factor to explain salary variations R-square 26%
2. USA Average data analyst salary USD 72K, and median USD 69K
3. CA Average data analyst salary USD 88K, and median USD 81K, at city grain San Jose, San Francisco are markedly higher within CA



4. Large firms don't necessarily pay more to data analysts
    - More hires/openings with higher salaries is more evident in these clusters. These high paying firms are typically unlisted /private with high ratings and have lower age (startups)
        - Mid-large firms (1k-5k employees, revenue USD 100-500 M)
        - Small firms (<50 employees, revenue USD < USD 5M
        - Small-medium firms (51-200 employees, unknown revenue)



5. Skill / experience – SQL, Excel; Python and technical degrees; MYSQL
    a. 60% of analyst jobs prefer SQL experience; 40% Excel skills, 30% prefer python experience, masters or computer science degrees; <4% positions require MYSQL but pay higher (+ USD 5K)
6. Regression Insights
    a. Job being in California makes its salary higher, irrespective of PHD degree preference of the opening.
    b. IT job doesn't necessarily pay higher; but being in California definitely does i.e., location is a bigger contributor to salary variations than sector
    c. New Jersey based firms (there are 74 firms HQ in NJ) that pay USD 5K higher
    d. "junior" descriptor in the title affects salaries USD 3K-8K lower. However there isnt any correlation to callout with other role titles.

Reference:

1. https://www.kaggle.com/andrewmvd/data-analyst-jobs
2. https://www.kaggle.com/yashvi/data-analyst-jobs-visualization