**Data Science Bootcamp**

**Project 2**

**Group members: Rich Kirschenheiter, Arielle Eagan, and Timothy Kelly**

***Proposal***

*Chosen topic and rationale:*

As future Data Science and STEM professionals, we are curious about the job market and earning potential in new roles available after we finish this Bootcamp.

Through Kaggle, we found a dataset of US-based Data Science and STEM Salaries available through Kaggle, which provides 62,000 salary records from top companies throughout the country. This dataset, which was provided in csv format, contains information such as company name, job title, location (city and state), compensation (total yearly compensation as well as base salary, bonus, stock grants), and demographics including education level, gender, and race. Looking at the dataset’s source, we found that the data was scraped from levels.fyi. Using this source, we found an API in JSON format that allowed us to pull more updated data, running only about 6 days behind current present day. Though this did mean that we no longer had access to We choose to use this API to ensure that our project and its visuals would have the most up-to-date data.

To add a second layer onto this geographic salary data, we also decided to use an API from Zillow to allow us an opportunity to add local housing rent and sale costs to our visualization(s) as appropriate. *If* we chose to include this, it may offer the user the opportunity to compare their potential salary in each state with their potential home rental or buying prices, which may influence their decision to accept an offered salary.

*Data and Metadata:*

* Data Science and STEM Salaries data set:
  + Original csv link via Kaggle: <https://www.kaggle.com/jackogozaly/data-science-and-stem-salaries>
  + Metadata:

Graphical user interface, text, application, email

Description automatically generated

* + Data source: <https://www.levels.fyi/?compare=Google,Facebook,Microsoft&track=Software%20Engineer>
  + API in JSON format: <https://www.levels.fyi/js/salaryData.json>
* Zillow dataset: <https://www.zillow.com/>
  + API: <https://www.zillow.com/howto/api/GetSearchResults.htm>

*Screenshots of relevant inspiring visualizations:*

Map

Description automatically generated with low confidence

<http://housing-markets.herokuapp.com/>

Map

Description automatically generated

<https://www.washingtonpost.com/technology/2019/07/23/ever-wonder-how-much-tech-workers-get-paid-your-town-this-map-might-have-answer/>

Teams

Description automatically generated

<https://www.visualcapitalist.com/us-states-top-tech-salaries-2021/>

Chart, scatter chart, bubble chart

Description automatically generated

<http://housing-markets.herokuapp.com/>

Chart, line chart

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

<https://www.visualcapitalist.com/us-states-top-tech-salaries-2021/>

*Sketch of final design:*

For our visualizations, we plan to create (1) an interactive map of the US with state-level salary so that the user can explore Data Science and STEM companies, job titles, and associated salaries across the country. We also plan to create (2) a bubble map to show the median base salary and median years of experience by state. Then we plan to create (3) an interactive line graph to show changing median salary for the top three most popular companies in the selected state on the map over the years (2017-current).

*Link to Github:*

<https://github.com/ArielleEagan/Project-2.git>