

Exploring Regional Disparities in Startup Funding and Growth

Team 6

Collaborators:

1. Gouri Dumale (G49564205) gouri.dumale@gwmail.gwu.edu
2. Nemi Makadia (G29362869) nemi.makadia@gwmail.gwu.edu
3. Bhanu sai praneeth Sarva (G46159306) bhanusaipraneeth.sarva@gwmail.gwu.edu
4. Smit Pancholi (G31443926) smitsnehal.pancholi@gwmail.gwu.edu

Description:

This dataset comprises approximately 54,000 rows and 39 columns, offering a diverse range of company-related information. It includes details such as company names, URLs, market sectors, geographical data (country, state, region, city), founding dates, and the timing of initial and final funding rounds. Additionally, the dataset encompasses various investment types, such as seed funding, venture capital, equity crowdfunding, undisclosed investments, convertible notes, debt financing, angel investments, grants, private equity, post-IPO equity, post-IPO debt, secondary market transactions, product crowdfunding, and rounds labeled A to H. The status of companies is classified into three categories: acquired, operating, or closed.

SMART Questions:

1. Can we predict the total funding amount for startups based on their location and the number of funding rounds? Specifically, how well do location and the number of funding rounds correlate with the funding total, and can we build a simple model to estimate funding amounts for startups?
2. What is the distribution of startup funding rounds across different regions and markets? Are there any notable differences in funding amounts between startups located in major cities versus smaller cities?
3. How does the presence of venture capital funding correlate with the location of startups, and what geographical patterns emerge regarding the availability of venture capital?

Modeling Methods:

We will employ linear regression analysis, specifically using the Ordinary Least Squares (OLS) method, to predict the accuracy of our model for the initial question. Additionally, we intend to conduct a classification analysis on the various features within the dataset to further enhance our understanding of the data.

GitHub repo: <https://github.com/smitpancholi313/DataMiningProject.git>

Dataset link: [StartUp Investments \(Crunchbase\) \(kaggle.com\)](https://www.kaggle.com/datasets/startup-investments)