Start ups in India

PATRICK LONG

Problem Description

Predict future Indian Start-up funding based on a number of factors such as:

-Amount funded

-Where the funding is coming from

-Which state the funding is going to

-Which industry is being funded

-When the start-ups are being funded

Client

Venture capitalist looking to invest in Indian Start-ups

Has capital for seed rounds, angel investing, or pre-seed rounds

Has knowledge of internet consumer services

Data Set

Two CSV files

First containing GDP of Indian states

Second containing all start-ups funded in India by Amount, City, Vertical, type of funding, etc.

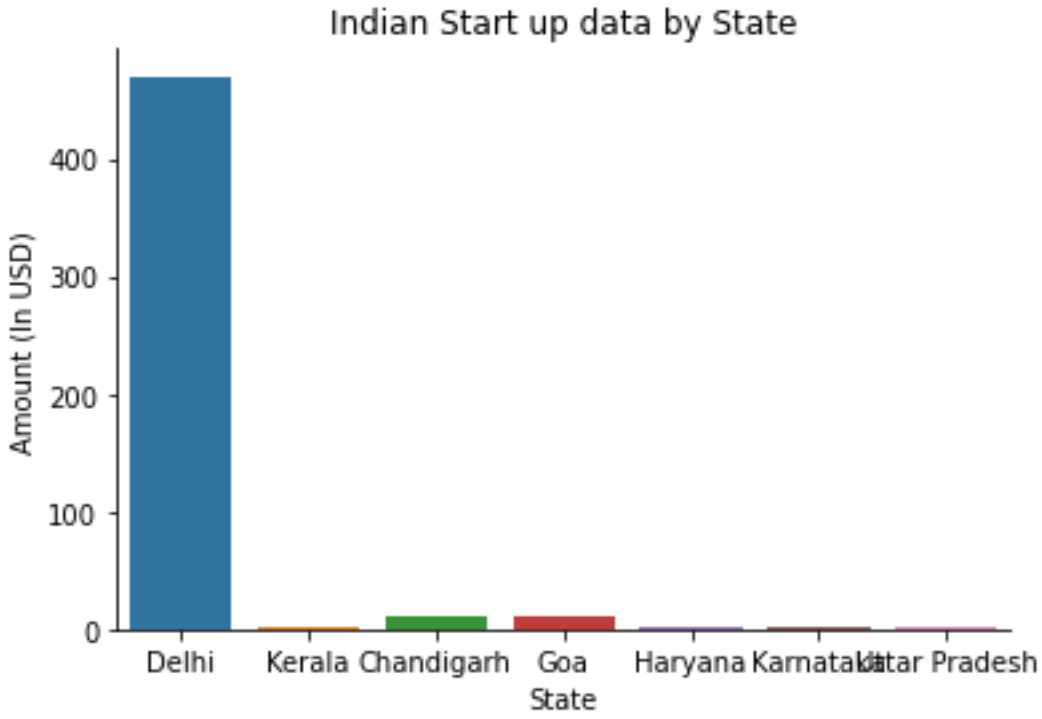
Data Cleaning and Wrangling

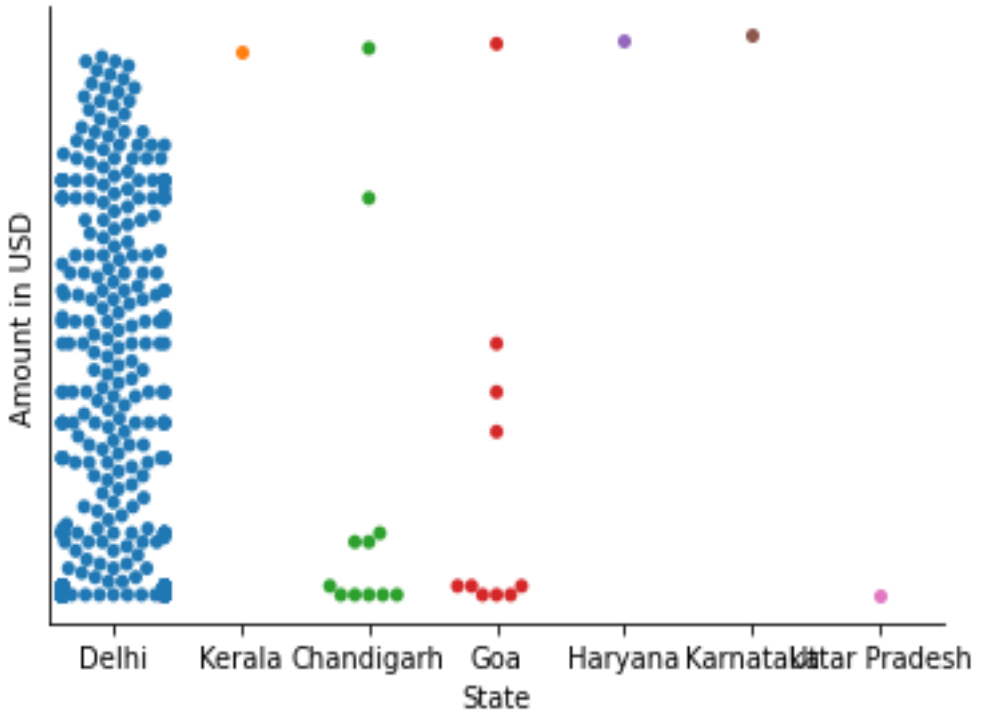
Mostly clean data sets

Cities were coded to which state they were in. There were multiple cities per state, so to see where GDP and start-ups were aligned.

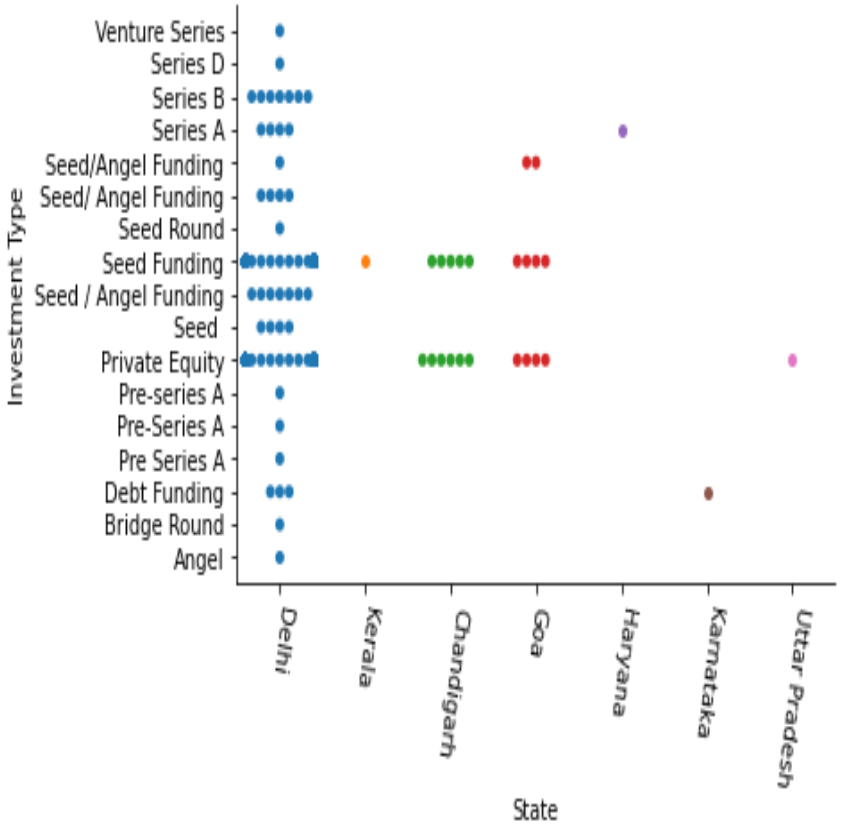
GDP was for each state were date from 2017 to 2019, so only captured start-up funding during that time.

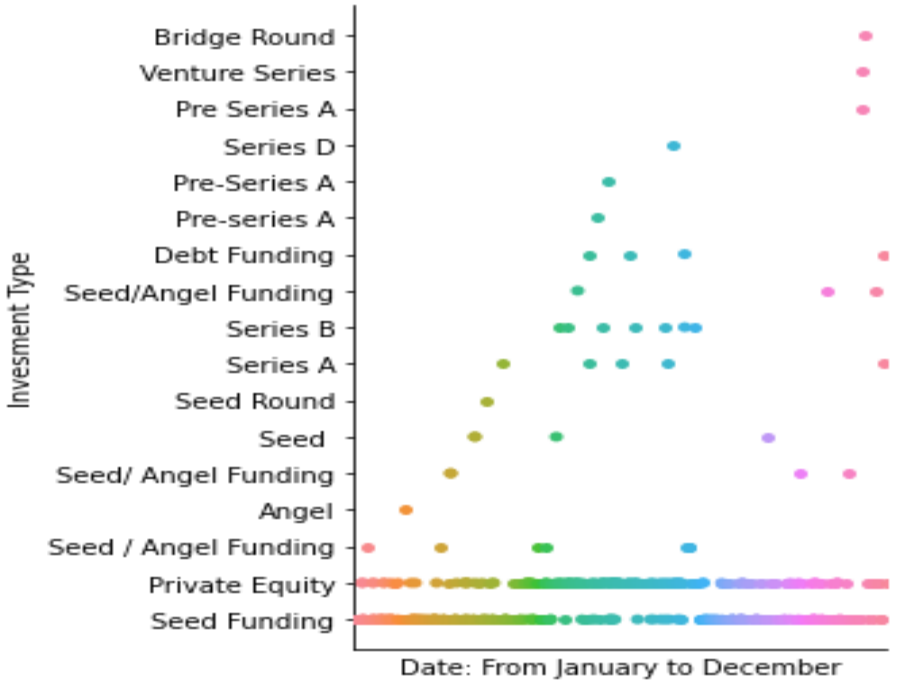
Exploratory Graphs



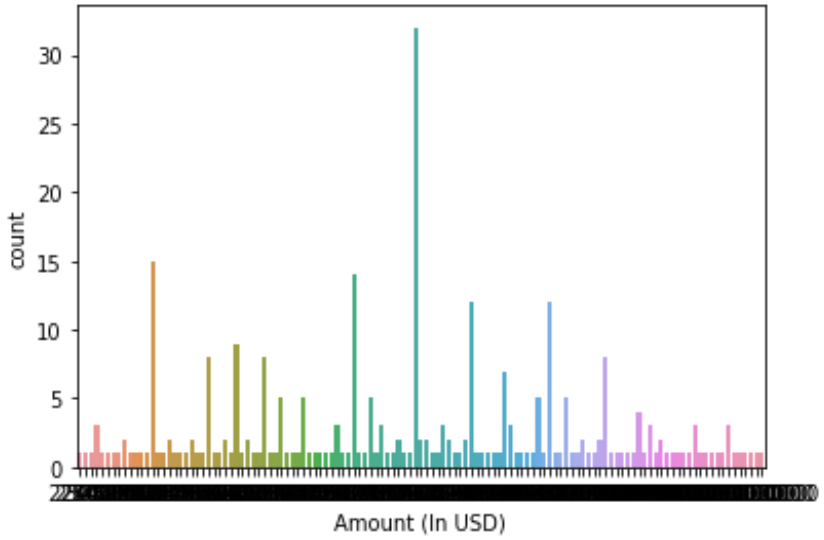


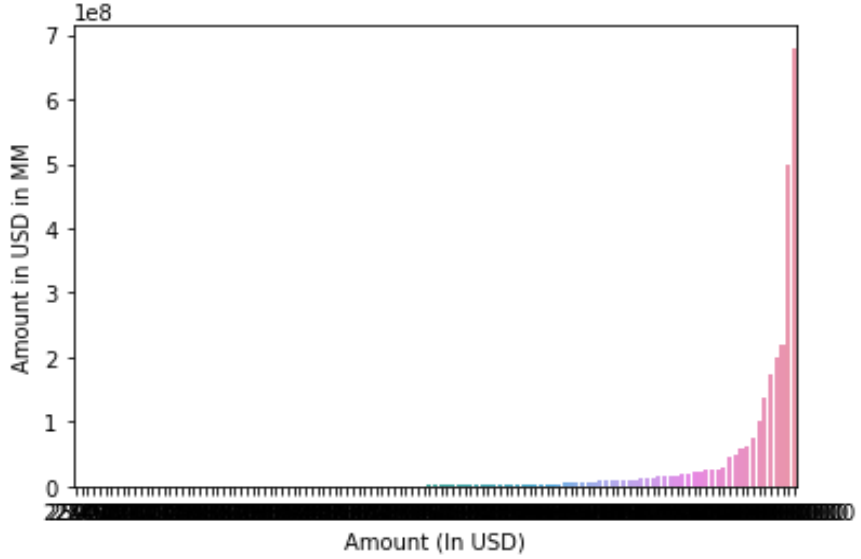
Exploratory Graphs





Exploratory Graphs





Results

Ultimately, trying to find and predict the best state in India to invest in is hard. Because so much of the start-up industry is in one state, Delhi, that is the de facto place to invest. My initial hypothesis was that Delhi would have the most startups with the service industry being the most numerous industry and private equity being the most popular funding method.

Those claims were quick to come to fruition, however the project was more revealing than that. What I was able to find is that yes, most startups are in Delhi there is a lot more to it. There are enough startups in other states to make it significant.

The investor that is looking for an extra edge would be prudent to start in a sector like Technology, in a state like Goa. Because there are still enough startups there to make researching and funding worthwhile, it will not be flooded with other investors like Delhi.

Why didn't the modeling work? This happens from time to time in Data Science. There are a few reasons that it could be. For this case it is data quality. That is not necessarily saying the data I had was poor. There just was not that much data. This can lead to those exceptionally low accuracy scores that were seen on each model. It could also be an example of overfitting. The data did not have much variance, so, the models were not able to 'learn' anything.

I would recommend any that is doing a project like this one. I.e., looking for startups in India to start with this project as the framework. Now that we know where most startups are coming form, how they are funded, and in what industry, I would search elsewhere. Look for government incentives to start and fund startups. Or look for other ways that might indicate where new startups might emerge. I recommend that because as we can see there is still room for growth outside of Delhi, but to know when and where that growth will come is the next challenge.

Recommendations

Start with Start-up search in Delhi.

Most start-ups are being funded through private equity and seed funding. Most likely to fund start-ups with that type of funding.

Other states mimic how Delhi is being funded, it is possible that contrarian funding could result in over-looked start-ups.