**Capstone Project Ideas – Sidhant Bhardwaj**

**Mentor- Eric Rynerson**

1. **Airbnb New User Booking**

Instead of waking to overlooked "Do not disturb" signs, [Airbnb](https://www.airbnb.com/) travelers find themselves rising with the birds in a whimsical treehouse, having their morning coffee on the deck of a houseboat, or cooking a shared regional breakfast with their hosts. New users on Airbnb can book a place to stay in 34,000+ cities across 190+ countries. By accurately predicting where a new user will book their first travel experience, Airbnb can share more personalized content with their community, decrease the average time to first booking, and better forecast demand.

The challenge is to predict in which country a new user will make his or her first booking. We have a list of user along with their demographics, web session records, and some summary statistics. We will predict which country a new user's first booking destination will be. All the users in this dataset are from the USA.

1. **Two Sigma Connect: Rental Listing Inquiries**

RentHop makes apartment search smarter by using data to sort rental listings by quality. But while looking for the perfect apartment is difficult enough, structuring and making sense of all available real estate data programmatically is even harder. The task is to predict the number of inquiries a new listing receives based on the listing’s creation date and other features. Doing so will help RentHop better handle fraud control, identify potential listing quality issues, and allow owners and agents to better understand renters’ needs and preferences. We will be able to predict how popular an apartment rental listing is based on the listing content like text description, photos, number of bedrooms, price, etc. The data comes from [renthop.com](https://www.renthop.com/), an apartment listing website. These apartments are located in New York City. The target variable, **Interest\_Level**, is defined by the number of inquiries a listing has in the duration that the listing was live on the site.

1. **Crunchbase Startup Companies Database**

Crunchbase is quickly becoming the dataset of record for the startup and venture capital communities. It can provide information on anything from what industries are hot ([biotech](https://modeanalytics.com/benn/reports/5dc33294f9c2)) to the potential effects of founder [experience](https://blog.modeanalytics.com/are-experienced-founders-better/) or [age](https://modeanalytics.com/benn/reports/211f54949c00). The dataset includes funding, investment, and acquisition data on over 40,000 companies. There are numerous questions that can be answered using the dataset such as are there characteristics of a company—industry, location, etc.—that differ by VC? Do some VCs typically invest together, while others rarely do so? Are companies raising more money earlier?