WHERE 2 START

TEAM NAME: 1-800-STARTUP
TEAM MEMBERS: Hansae Lee, Eun Hye (Grace) Oh, Iris Nayki, Matthew Davis, Tianxing Liu, Yuting Xu

DATA SCIENCE QUESTION(S) & HYPOTHESIS:
Question: Does location and industry contribute to the overall success (defined by the total funding and market value) of a startup more than other factors?
Hypothesis: The location and industry of startups are the main contributors to the overall success of a startup compared to other factors. We believe so because location is often correlated with different amount of business opportunities, and industry often dictates the profit margins and market.

BACKGROUND:

Given the continual rise of startups in various locations and their diverse industries, there is rich data regarding their funds and overall performance. We are curious as to what factors contribute most to the success of startups, such as their specific industry and location and how much and in what ways they may contribute to their funding and market value. We hope to track trends and commonalities among a great number and wide range of startups to answer our questions and approach our hypothesis.

A similar project on this topic is Startup World where they used Crunchbase data, Loopnet commercial real estate prices, and coworking locations in order to find "the most lucrative cities for startups and coworking places today". After extensive research and analysis, Diana Kontsevaia concluded that the best place for startups to work is Milwaukee, Wisconsin since it has a high average funding.

ETHICAL CONSIDERATIONS:
For the ethical considerations, we referred to Deon's Ethics Checklist:
Data Collection:

The data we are using is from a database called Kaggle, which provides datasets to the public. Their terms and services do not require their users to ask any consents to use their data unless the user is underage. The dataset itself only contains quantitative values that state simple facts, such as founding dates of companies or amount of funding, for all the startups in the region, so it is free from collection/source bias. All the data are about each startup company, so there is no exposure of personally identifiable information.

Data Storage:

We will store our datasets in our Google drive which only our team members have access to. Any unnecessary data will be removed from the storage during the process. After submission of this project, all the remaining data in the Google drive will be deleted.

Analysis:

We are measuring the performance of startups based on factors such as investments, lifespan and profits. However, it's hard to quantify the success of a startup precisely.

DATA:
In order to answer our research question, we reviewed Kaggle, an online community for data scientists to share datasets and their findings. After some time, we found the dataset, "Startups and Investors from Crunchbase". This dataset includes over 17,000 startup companies with categories that include the company's industry, the total funding received, the location of the company, when the startup was founded, and the current status of the company. The data on Kaggle was last updated a year ago and was scraped from Crunchbase, a platform meant for public and private companies.

Team Expectations Agreement

Group members will show up to most meetings and complete their assigned tasks according to the team's internal project deadline. Members will communicate inability to attend meetings as well as their work progress, if needed by the team. All members will meet their own deadlines for their parts, and every member will equally contribute to the project. When the team comes up with any decisions, all members will agree to it.

Project Timeline Proposal

Include actual dates and times for due dates and meetings below, not just what week they'll be completed

	Draft Text?	Write Code?	Proposed due date	Discuss at team meeting	Edit?
Initial team meeting	NA	NA	NA	04/11	NA
Background Research	Grace, Iris	NA	04/21	04/18	Yuting
Question & Hypothesis	Grace, Iris, Tian, Matthew, Yuting, Hansae	NA	04/21	04/18	NA
Ethical Considerations	Hansae, Yuting	NA	04/21	04/18	Iris
Dataset	Iris	Iris	04/21	04/18	Yuting
Descriptive	Matthew	Matthew	05/02	05/09	Grace
Exploratory	Yuting	Yuting	05/02	05/09	Tian
Analysis - Part I	Iris	Iris	05/09	05/16	Matthew
Analysis - Part II	Tian	Tian	05/09	05/16	Yuting
Analysis - Part III	Hansae	Hansae	05/09	05/16	Tian
Summarize Results	Yuting	NA	05/16	05/23	Grace
Conclusions	Grace	NA	05/16	05/23	Matthew