- **Team Name:** Powder
- **Group Members:** Alex Ojemann and Andrew Floyd

- **Problem Space**
  - **What is the problem you would like to solve? Try to describe it in terms of ML.**
    - We would like to predict qualities of the snowpack, like volume or moisture content in terms of Snow Water Equivalent (SWE). It's a regression problem because we're predicting a numeric variable rather than a category and it's a continuous numeric variable.

  - **Why is this an important and interesting problem? What triggered the idea?**
    - We both enjoy skiing and snowboarding, and it's useful when deciding whether or not to drive to the mountain whether there's a good amount of snow. Knowing this in advance allows us and others to make better decisions.
    - Snow Water Equivalent is a measure of the water content of snow. It can be thought of as the amount of water that would be produced by melting all of the snowpack and collecting it together as liquid water. This can vary widely – wet, slushy snow, while less voluminous, can contain quite a lot of water. As a metric, it is used by hydrologists, policy-makers, and others to estimate the availability of water resources, an important input in flood planning, reservoir management, and much more.

  - **A description of why it would be useful to use / develop machine learning to solve this problem.**
    - Machine learning algorithms could allow us to model the snowpack depth and SWE based on many different factors that we may not consider or be able to understand the impact of when predicting the amount of snow on our own. For example, our data set includes humidity information. Skiers may not know that higher humidity tends to result in lower snowpack because it allows the snow to melt more easily, or they may know that but not be able to contextualize how much influence this has. A machine learning algorithm could do so better than humans by analyzing trends in a large set of data.

- **Data/Data Plan**
  - **Description of Data**
    - Our data is derived from a variety of sources, generally publicly accessible through state, national, and international sources. Much of it comes from satellite imagery and Lidar data or from remote sensing stations located throughout the US and, in particular, western states.

  - **Potential Data Sources**
    - **NASA / JPL Airborne Snow Observatory (ASO)**
      - https://nsidc.org/data/aso/data
    - **NOAA High-Resolution Rapid Refresh (HRRR) Meteorological**
      - https://rapidrefresh.noaa.gov/hrrr/
    - **MODIS and VIIRS Snow and Ice Global Mapping Product**
      - Terra MOD10A1 (Snow Cover Daily L3 Global 500m SIN Grid)
      - Aqua MYD10A1
      - https://modis-snow-ice.gsfc.nasa.gov/
    - **NRCS SNOTEL Sensor Data**
      - Daily, Hourly, Weekly, …
      - https://www.nrcs.usda.gov/wps/portal/wcc/home/
    - **Copernicus DEM Digital Surface Model**
      - EEA-10, GLO-30 and GLO-90
      - https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model
    - **California Data Exchange Center (CDEC)**
      - https://cdec.water.ca.gov/queryTools.html

  - **What are some of the interesting or critical features you have?**
    - Global Positioning Data (e.g. Latitude, Longitude, Altitude)
    - Elevation Gradients
    - Regional Climate Zone(s)
    - Meteorological/Climate Data
      - Temperature
      - Relative / Specific Humidity
      - Barometric Pressure
      - Precipitation Totals
    - Average Snowpack Depth
    - Snow Cover / Fractional Snow Cover
    - Daily Snow Albedo

- **Are there any features you plan to exclude?**
    - None that we know of yet, we will discover which features are useful and not as we continue to explore the data.
- **Outcome variable(s) or other expected results**
    - Snowpack depth and Snow Water Equivalent (SWE)
    - Snow Quality
    - Estimated risk of drought / flood / ‘water (in)security’
- **Is it supervised or unsupervised?**
    - We might lean towards a supervised learning model, and rely on labels that can capture our outcome variables into general categories (eg. high/med/low drought risk, great/okay/rowdy snow, etc). We may establish reasonable thresholds for snowpack depth or for snow water equivalent, about some key date(s), as proxies for identifying difficult-to-observe outcomes, like ‘great snow’ or ‘drought risk’.
- **Approximately (or exactly) how many samples do we have?**
    - Satellite imagery data is available for large swaths of the country and in particular the western United States. Resolution of the data varies from as fine as tens to hundreds of meters upwards of 1-3 km or more. It may prove necessary to combine or otherwise reshape data points into a suitable format for our purposes.
- **How we will get the data.**
    - While we have identified several sources of data, it is quite possible some of it may not prove useful. Alternatively, we will need to think more about the most efficient way of pulling down the data, storing it, and making it accessible to our model.

Interesting Links and Other Sources of Information:
https://github.com/drivendataorg/snowcast-showdown
https://github.com/nsidc/NSIDC-Data-Tutorials
https://github.com/hanis-z/Snow-water-equivalent

Example Data:
https://github.com/hanis-z/Snow-water-equivalent/tree/main/data