Principles

- Document everything (or just enough)
- Everything in Git (.gitignore, environment)
- Automate or Code as much as possible
- 'Critical' decisions should be discussed as a group
- Meet daily to keep teammates informed about status (What did you do, what will you do, any blockers)
- Pairing on interesting and painful topics is encouraged
- Kanban

Data Sources and Preparation

Create Output Master Table with "empty" time windows

Understand all data sources

Look for (and understand) potential additional data sources

EDA

Under which circumstances do accidents happen (day, time of day, month, road segment, coordinates, weather condition, speed etc.)?

Accident prediction

How to do time?

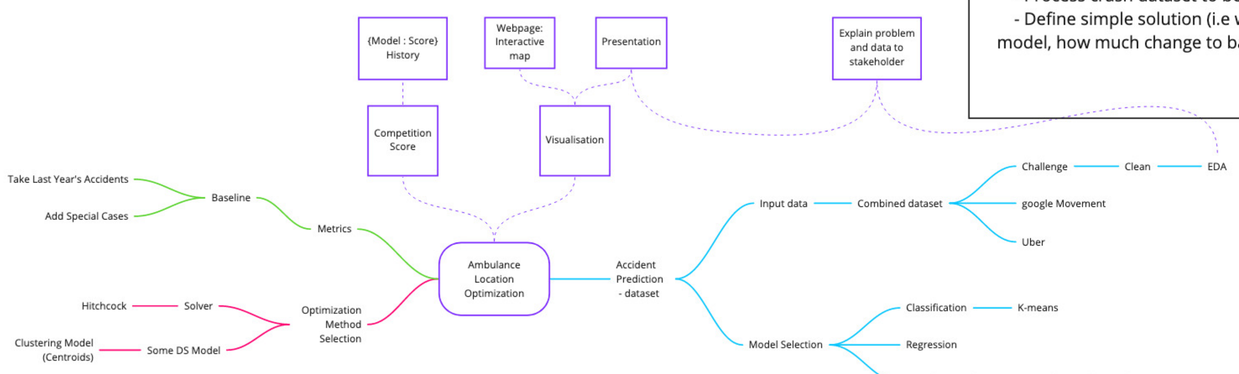
How to do location?

Classification or Regression?

Week One objective:
Submit Solution to Zindi (as automated as possible)
Present Introduction to problem, solution method and baseline

Milestone: 1st Solution Submitted
Milestone: Consideration Set of clustering algorithms identified with relevant pros and cons
Milestone: Consideration Set of prediction algorithms identified...
Milestone: Slides / Presentation of Problem as introduction

Tasks:
- Create way to store submission and resulting score
- Understand submission format
- Understand submission process
- Script written to generate submission from prepared data / model output
- Create model to place ambulances based on input data and problem approach
- Process crash dataset to be used in first output model
- Define simple solution (i.e what is input, what is baseline model, how much change to baseline do we want)



Parking Lot for Extensions

Evaluation function for advanced project - have a better metric that takes time or distance as a measurement and "blocks" an ambulance for a certain amount of time after it gets used. (Latter point might lead to re-calculation of optimal locations after each accident?)
Idea from Tereza: Probability of someone dying (could increase over time)
Building on that - have a classification (sort of a histogram?) of how many people were attend to in: less than 5, less than 10, less than 15 minutes etc.

Heirarchy of Metrics:
Euclidean distance (Challenge Metric)
Road distance
Time (based on avg segment speed)
Golden Hour Threshold to measure deaths

Extension: Dealing with multiple accidents in time period
How to deal with multiple accidents in a time window? e.g. run model with n-1 nodes to reallocate waiting zone to ambulances:

For how long are ambulances taken out of the pool when a close accident occurs

How many times no ambulance available (i.e. accidents in time period greater than no. of

Problem Definition:

Ambulances are a limited resource
Accidents are difficult to predict
time for an ambulance to reach an accident site is critical to outcome. (every second counts - golden hour)

Positioning ambulances in a city determines the time needed.
Current methods for positioning rely on local knowledge in individuals heads.
Placement of each Ambulance is not coordinated to optimize the entire system.