

Week-01 (Learn about Capstone Basics)

* Build portfolio.

Next plan :

→ LinkedIn post.

→ Make portfolio using Kaggle.

* Case study best approaches :

① Make sure the case study answer the question being ask.

* Case study: How does a Bike-share Navigate steady success?

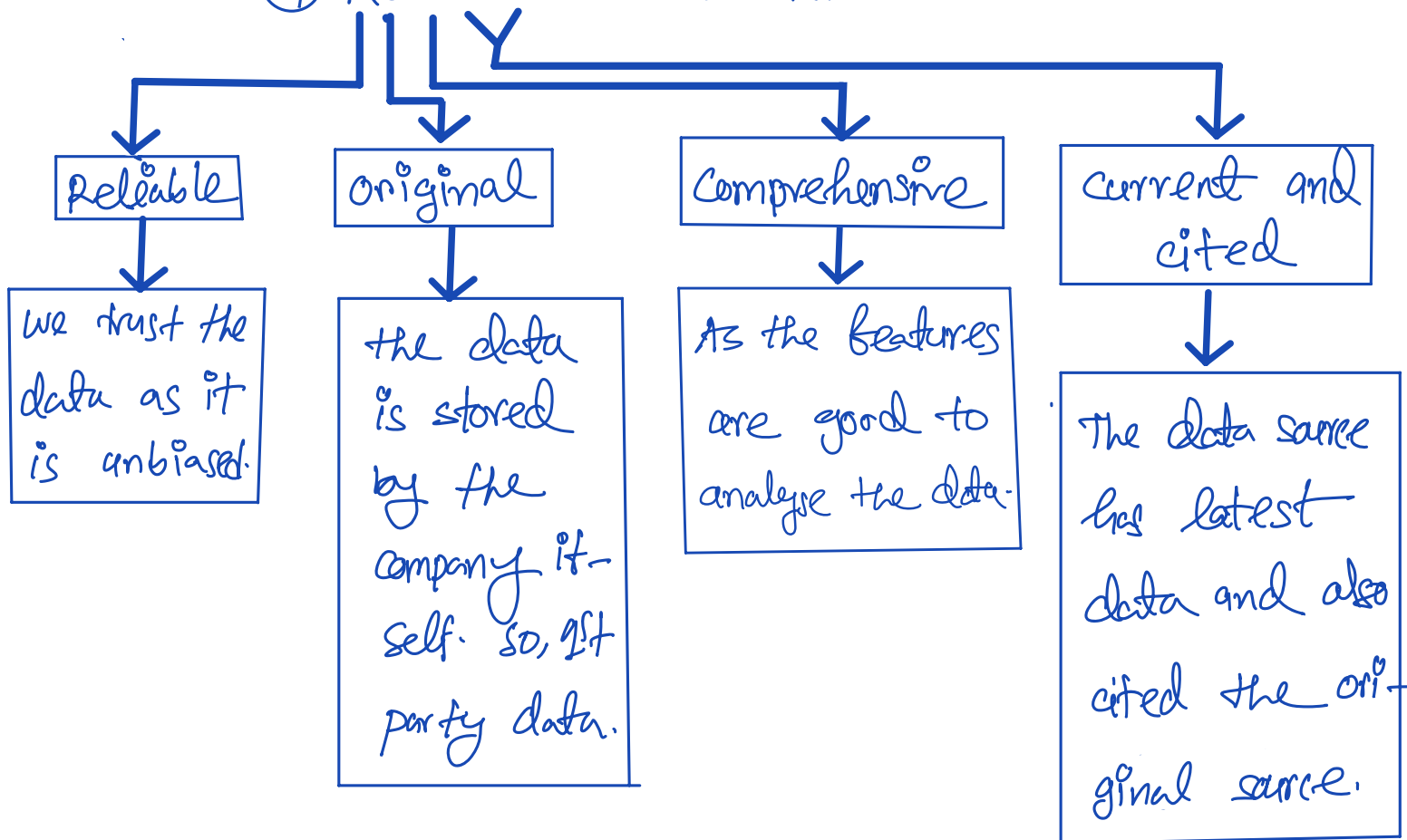
→ Goal of the analysis :

1. Analyse how different types of membership vary in the dataset.

2. Findout why its important to convert casual riders to anual member.

→ About the data:

- ① Data is located in an open-source web.
- ② Data is stored into csv files, where it's available from 2013 to 2023 (March). For the analysis purpose I use the data from 2022.
- ③ No bias found for the data.
- ④ ROCCC characteristics.



→ checking the integrity of the data:

- ① accuracy: Data is accurate.
- ② Though there are some null/empty cell but overall data are complete.
- ③ Data are consistent.
- ④ As the data provided by the company / 1st party data. so, trustworthiness achieved.

* Problem in the dataset:

- ① Most null values are in station info columns.

start-station-id or start-station-name = 833 K

end-station-id or end-station-name = 892 K.

end-lat or end-lng = 5.8 K

so, In all 4 column has null values for \approx 5.8 K

↓
(contains: end-station-info)

② Lat-long data is confusing because of precision.

so, to fix it and make same lat/long for a specific station.

① Only took 5 decimal point.

② Took median of these point

③ And assign it to the lat/long for all of the station.

* Processing :

From the start and end time we need to calculate the length of the trip and the day of the week.