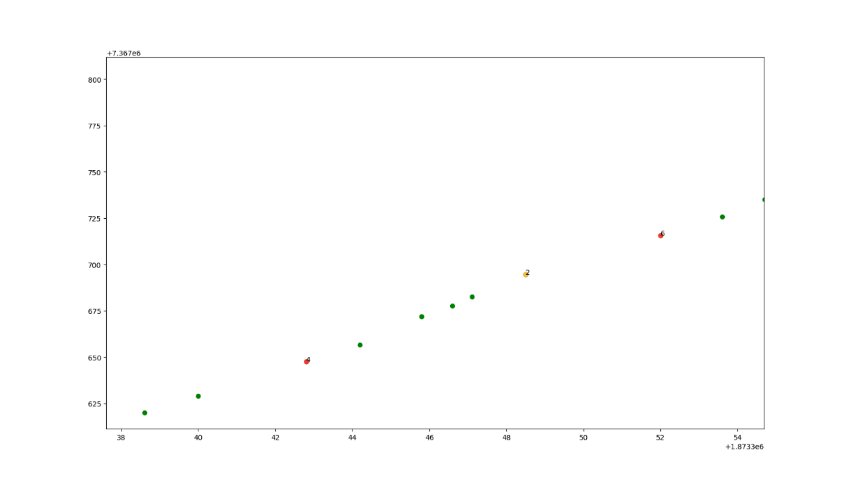
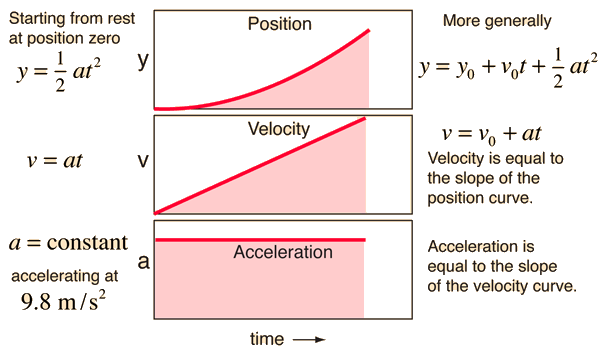
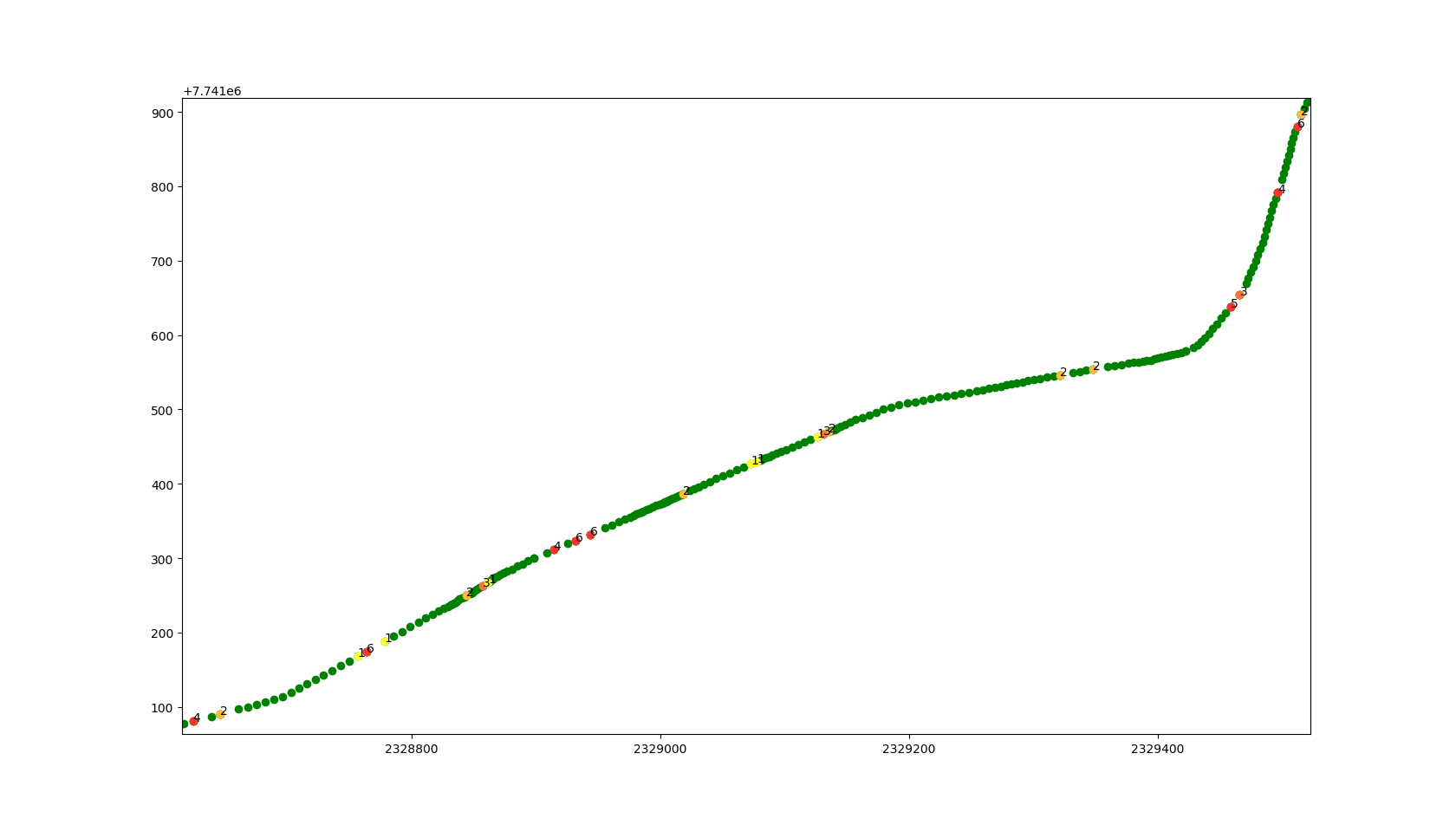
Simple GPS anomaly and missing data detection in real dataset

**-read all the new data:**

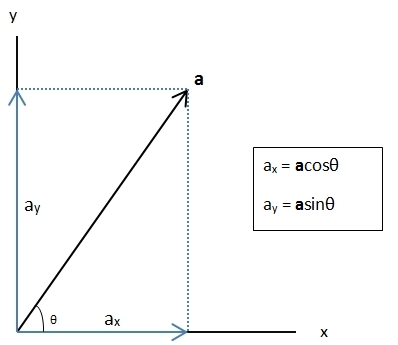
Had to optimize spark data reading and curated data => RESULTED 4 GB OF CSV DATA

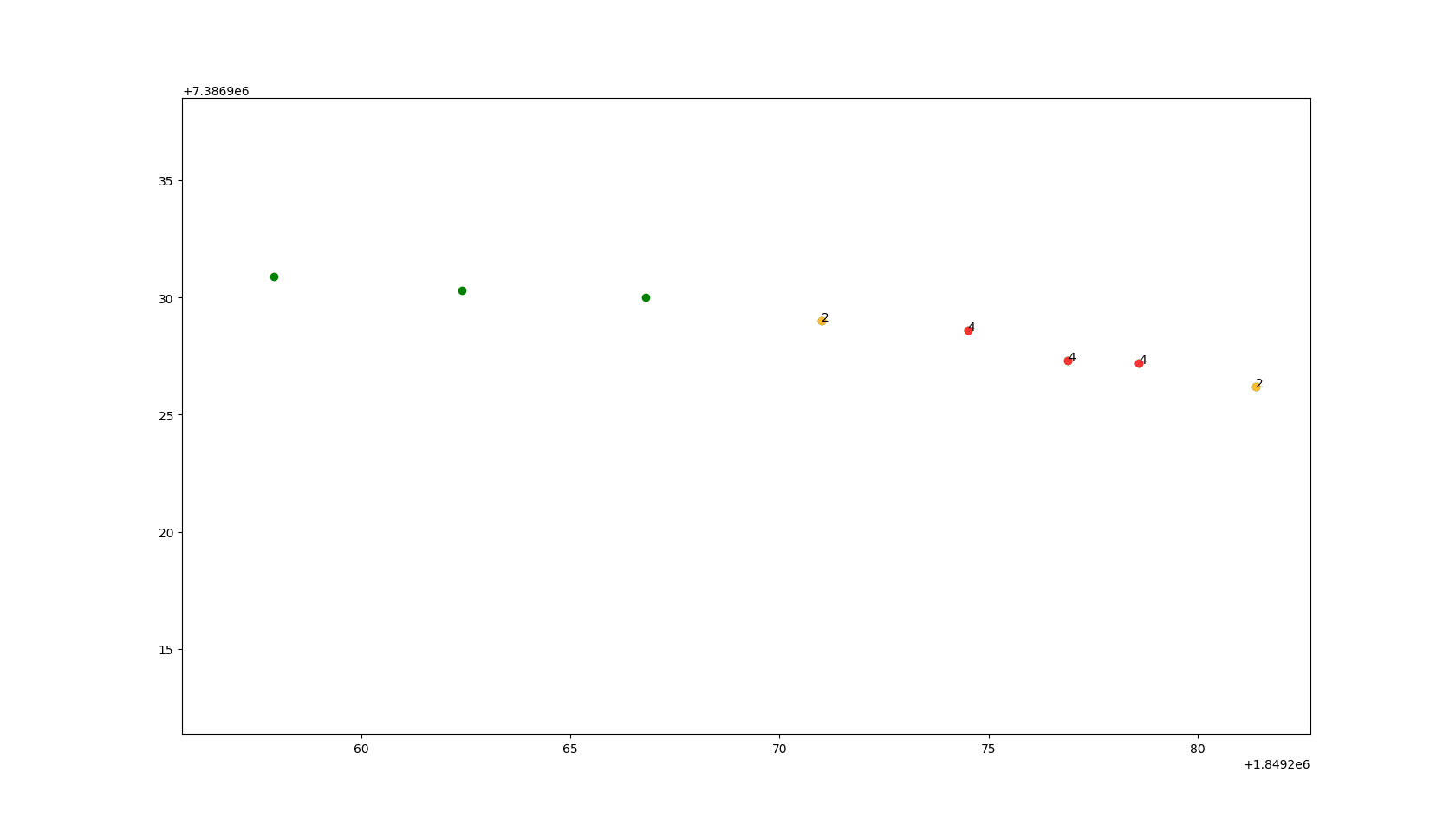
**-mathematical model :**

****->first based on acceleration differences

 ****

->second based on velocity based on angle differences

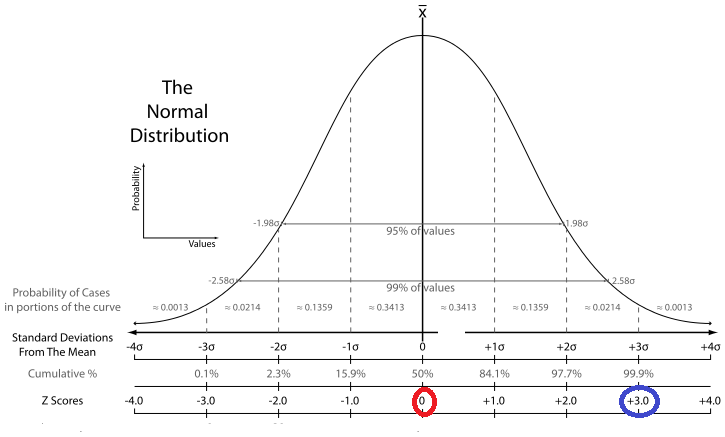


****

Gps points -> distances ---> speed --->mean speed----------🡪

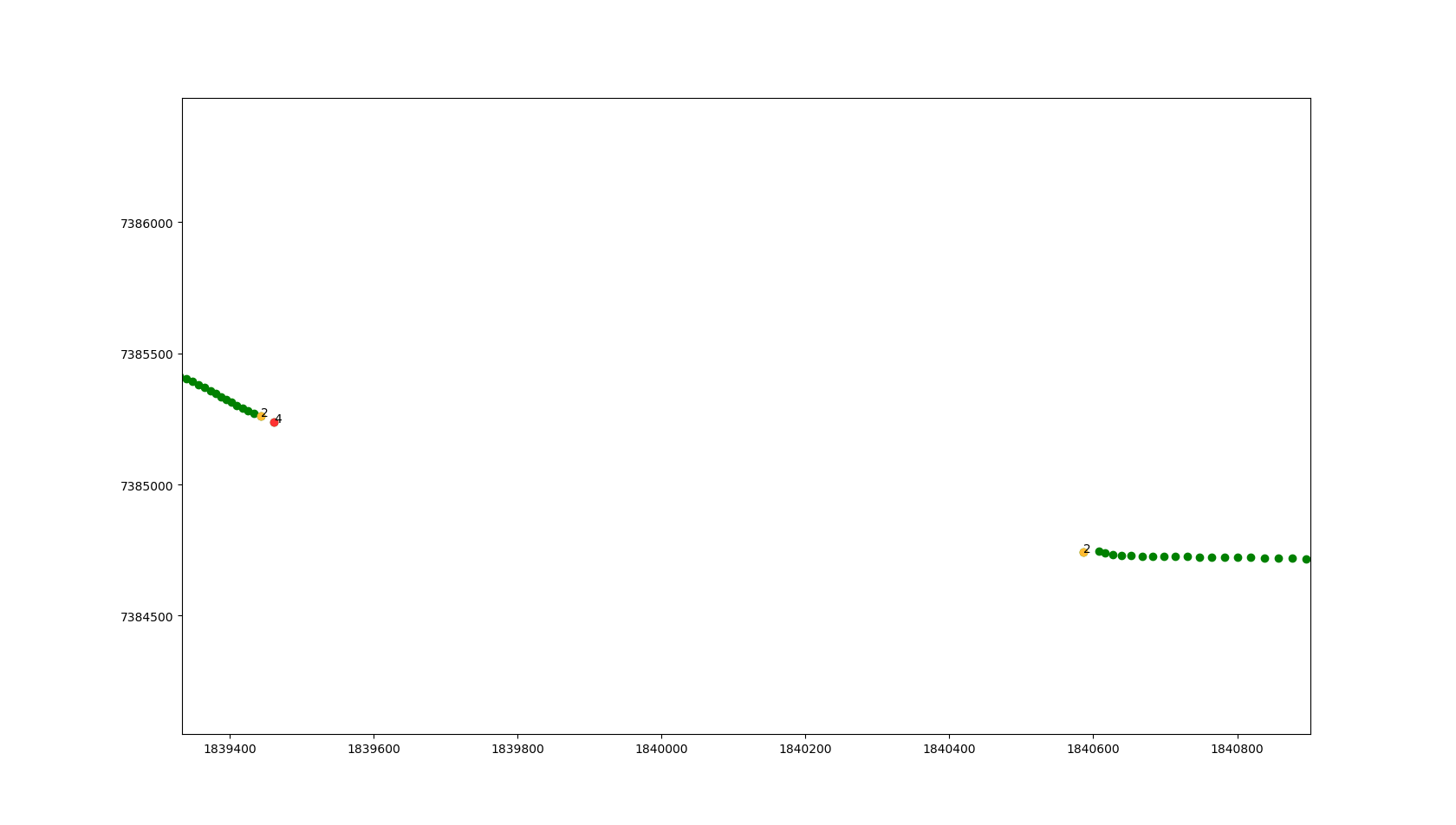
Time ---> ---->speed distribution-🡪Z-SCORE

---->standard deviation🡪

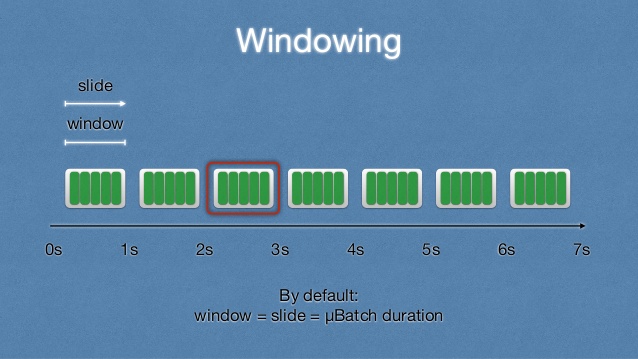


Obs: quantity of input data and of the Z-Score have a big impact on the results.

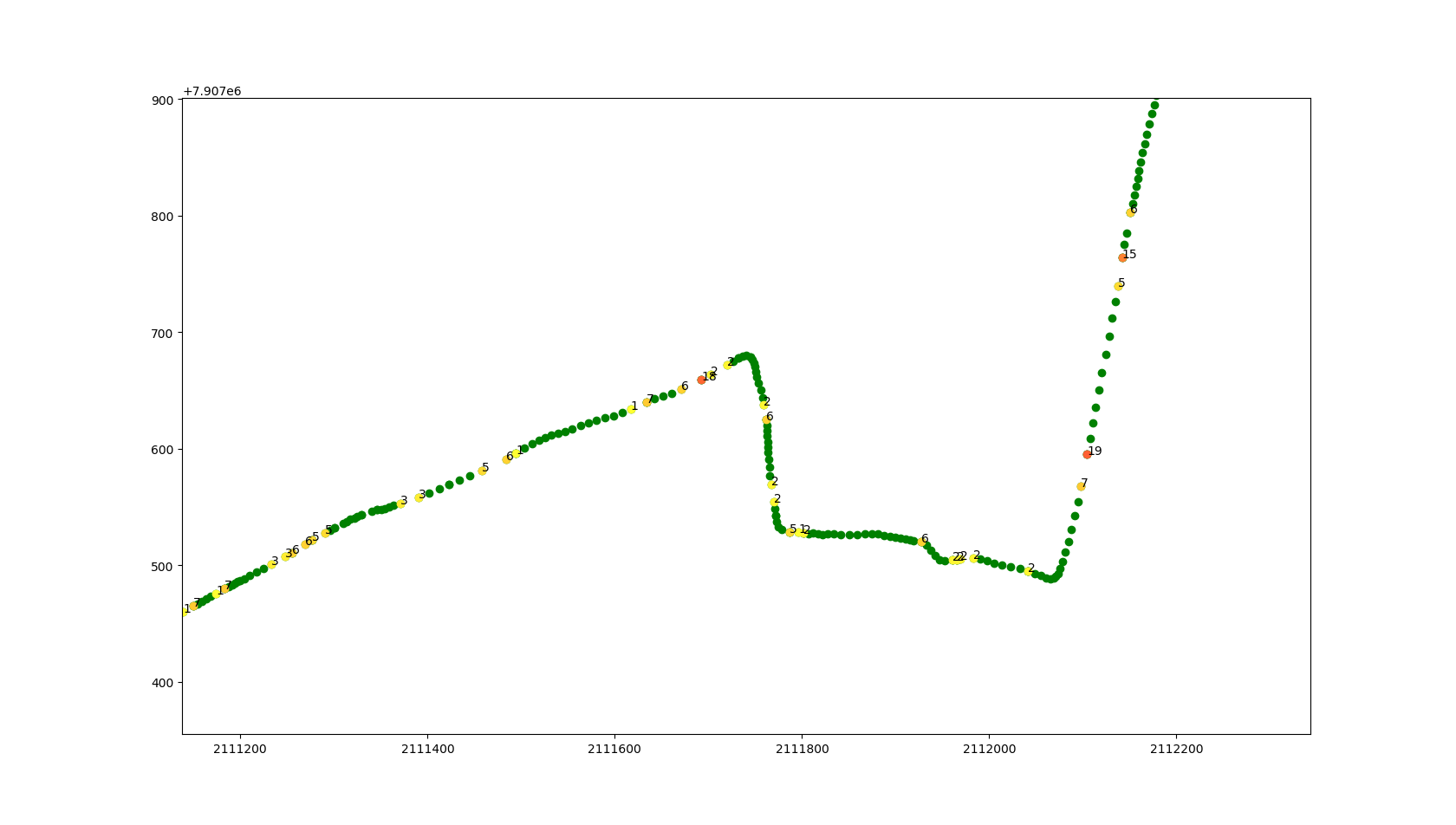
Ex: extreme case :

****

Combine both methods and window sliding for best results



**-mathematical model with graphic visualization:**  numbers represent the gravity of the anomaly

****

**-mathematical model export results:**

In two CSV :

1. with all the classified data as Mediocre to Terrible ( 1-24 )

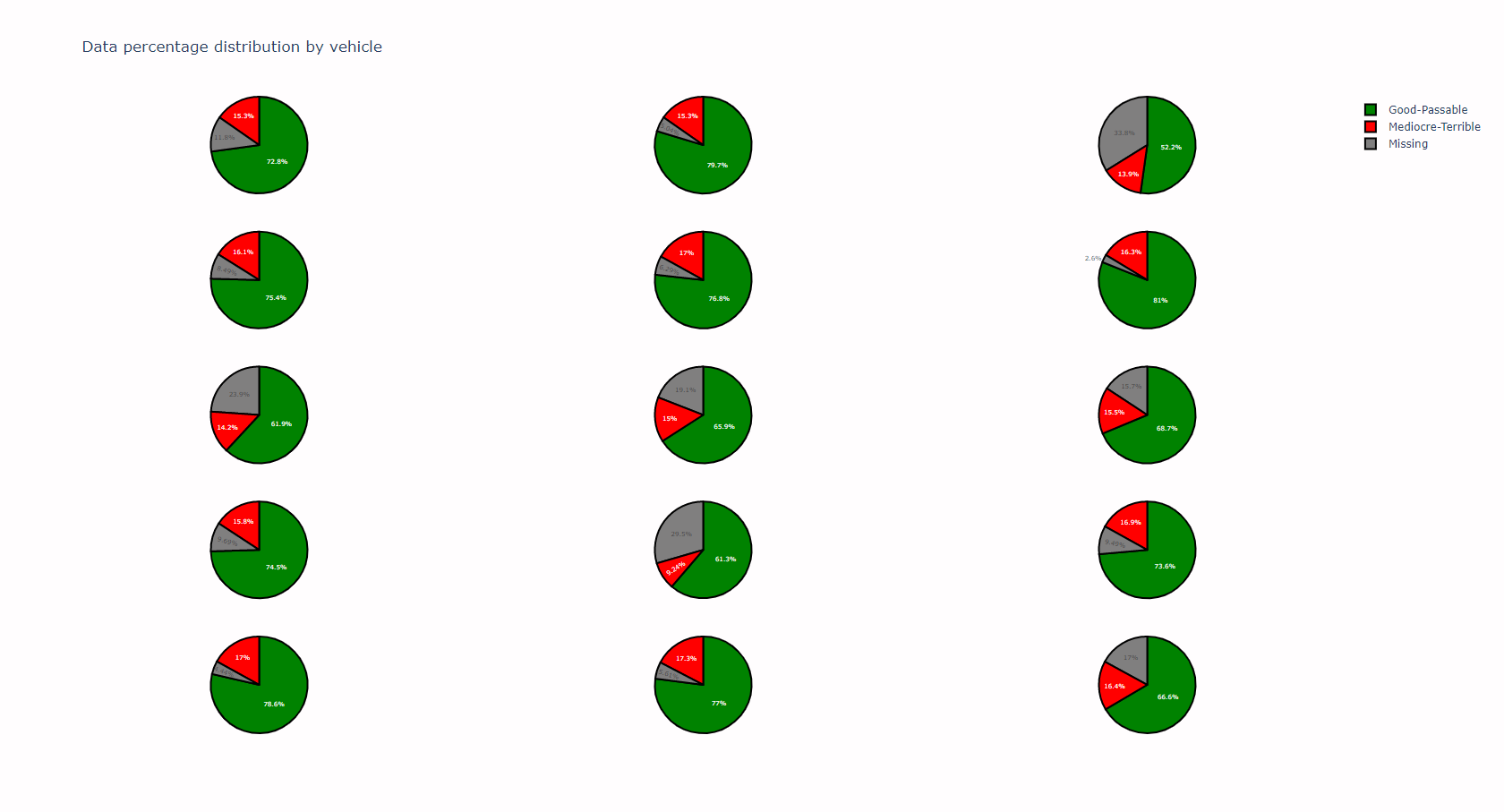
41.468765,-81.769044,6,2019-02-16 05:00:12.001000  
  
41.468813,-81.769082,9,2019-02-16 05:00:11  
  
41.463521,-81.769006,8,2019-02-16 04:59:30.001000

2.with all the inexistent data

None,None,2019-02-01 15:07:57  
  
None,None,2019-02-01 15:07:58.001000

**-Visualization using pie charts**

Counts the data of every category and return a report



**Simple visual representation**

SPARK READ

**TIME,LAT,LONG**

**TEST**

**LIST**

**LAT,LONG, TIME,RANG**

**TRAIN**

**TRAIN**

**LAT, LONG, TIME, SPEED, ACC, VECTOR RELATIVE VELOCITY, RANG**

**LAT,LONG, TIME,RANG**

**DF**

**DF**

**DICT**

GEN

ML MODEL

GEN

MATHEMATICAL MODEL