**1- What validation of data you will do:**

* Validate data types of fields:

1. Decimal: (start\_station\_latitude, start\_station\_longitude,end\_station\_latitude,end\_station\_longitude)
2. Integer: (duration\_sec,start\_station\_id,end\_station\_id,bike\_id)
3. String: (start\_station\_name,end\_station\_name)
4. Time: (start\_time,end\_time)
5. Date.Year(member\_birth\_year)
6. DDL: (bike\_share\_for\_all\_trip {Yes,No}, member\_gender {Female,Male,Other}, user\_type {Customer,Subscriber})-->String

* Validate that there are no duplications of records
* Decide about Null values depending on the application that will use this data, ex: to drop or to fill them with the median(temperature related application)

**2- Assume all validation failed, how are you going to fix the data set**

This depends on the business use of the application that will use this data:

* Clean data OR
* Drop corrupted records OR
* Cast values (ex: date time to time)

**3- What questions are you going to ask for this dataset?**

* What's the average of trip duration?
* What's the count of trips for each user type?
* What's the count of bike trips by each member gender?
* What's the relationship between bike trip duration and member age?
* What's the relationship between bike trip duration and member gender?
* What's the relationship between bike trip duration and user type?

**PS** Visuals were done on Power BI to answer these questions.