Hi Mark(Product Leader),

I have reviewed the data and done some research on it and would like to share some thoughts about the data and its quality based on some of the questions.

Before are some of my answers to key questions

Questions on data:
What exactly we want to find out from the data
What are the key Business insights that this data can help
What outcome from analysis would you deem as a success
What Standard KPI will you use that can help business
Where is or will this data coming from?
What scales would you apply for this dataset

Discover data quality issues:
The data quality issues are discovered by checking some of the key things below
Inconsistent data - Same records exist multiple times in a database.
Poorly defined data - Data's are sectioned in a wrong category. For example some of the dates that mentioned here are not in data format and with incorrect data type.
Data columns are not consistent with names and it is really difficult to identify the nature of the data with the column name

Resolve data Quality issues:
Fix in the source system by cleaning up the original source of data
Fixing of duplicate records
Identifying the null values and replace them with some valid values
Identify the outliers and remove them if necessary for analysis
Fix the incomplete formats and replace them with correct format
Normalize the measurement units across the data to get the appropriate results

Optimizing the data Assets:
Data Profiling - Examine the data defects. It also analyses the uniqueness of data. This can be achieved by using data mining tools for data quality purposes
Data Normalization - As data is collected from various sources it may have a variety of spelling options.It also helps remove redundancy in data.
Prevent and fix the duplicate records
Make sure data is secure

Performance and Scaling concern in production environment:
Accuracy with the data on how well this data is measured in the real world
Timeliness of data like how recently the event this data represents
Consistency with the data on multiple versions
Validity on how these data is loaded with the correct formats
Best,
Srinath