**Run instructions. How should we run your code?**

Requirements:

* Python 2.7.11: https://www.python.org/downloads/release/python-2711/

Python libraries:

* + Pandas (<https://pypi.python.org/pypi/pandas> )
  + Xlrd (<https://pypi.python.org/pypi/xlrd> )

Instructions:

* Download contents of the folder “Package” to your machine.
* Open command prompt
* Browse to the “Package” folder.
* Run Command: C:\Python27\python.exe Normalizing.py
* Output:
  + OutputError.xlsx: Has Error entries based on assumption.
  + OutputNormal.xlsx: Has normalized data used to derive the various trends.

***NOTE: AnalysisCharts.xlsx in the package folder contains charts created manually using the Normalized data generated from the program.***

**Brief write up on assumptions you have made when solving the problem.**

* Negative entries in the “Value” column are Errors.
* Time column represent “Timestamps” (noted increments of 60\*n units)
* The raw data is

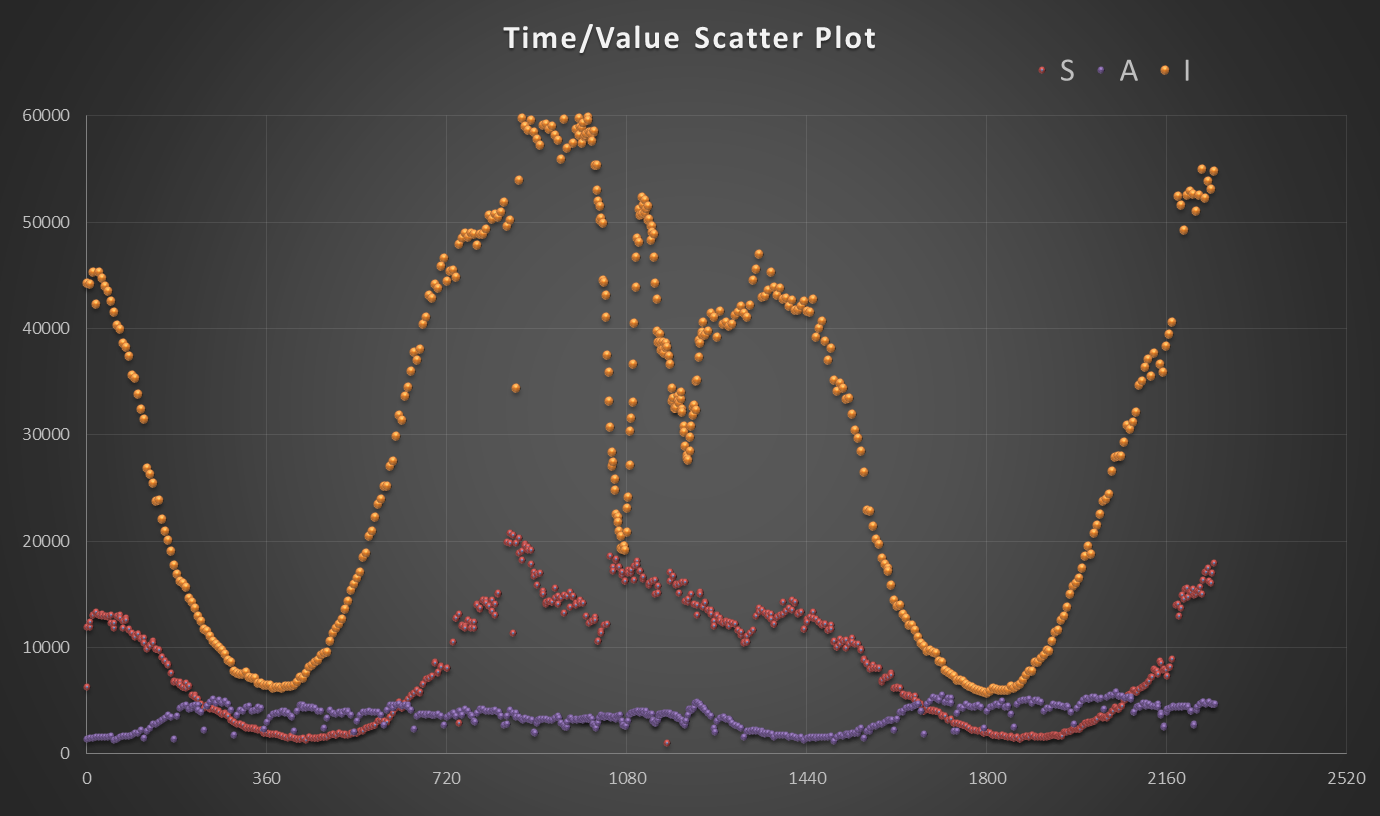
**What were your findings?**

**Observations:**

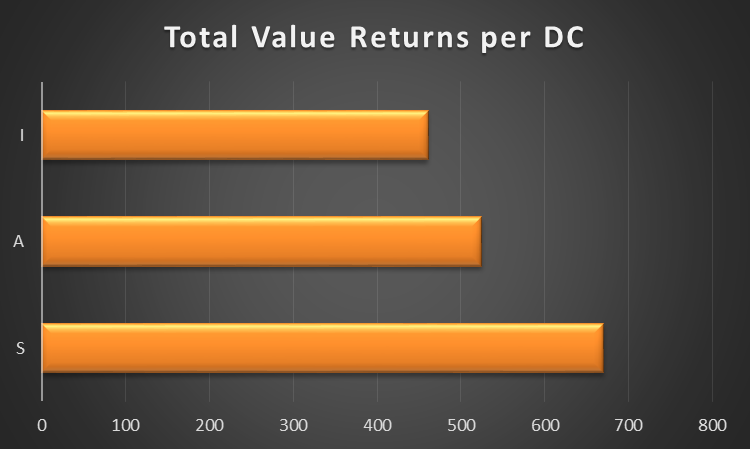
* 22: Value Errors
* 937: Unique Time Entries excluding errors
* DC-S has 670 values, A has 525 and I has 461.

**Trends:**

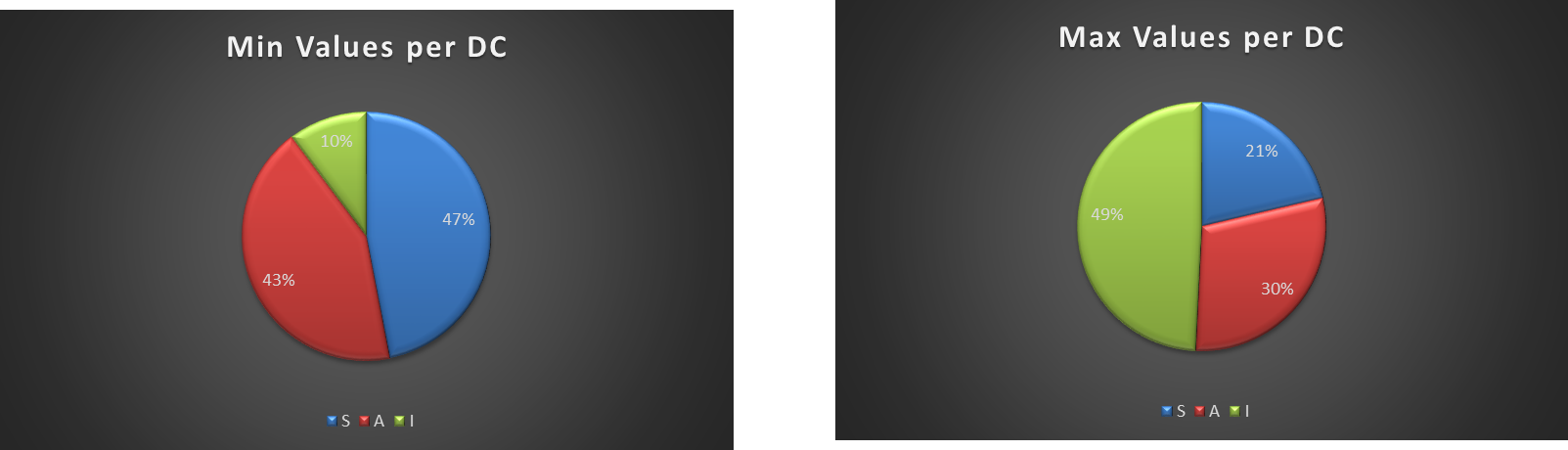
1. DC-I tends to have highest values, DC-A tends to have lower values.
2. The space between the troughs (for data centers I & S) represents 1440 time units (a day)
3. For a specific time window DC-S has lower values.



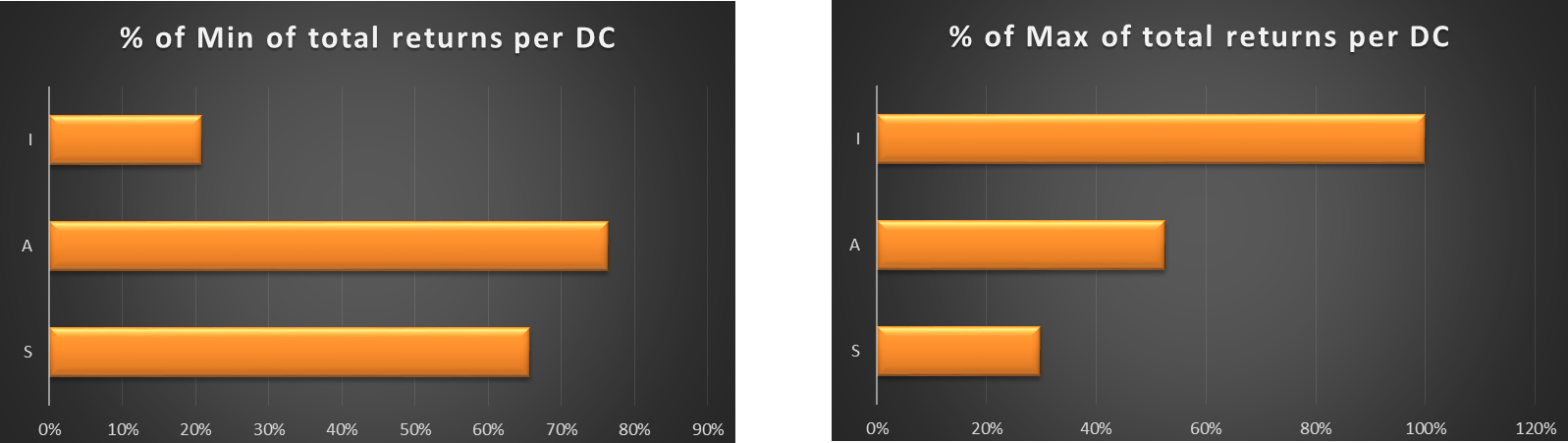
**Data center S has the most values, I the least.**

****

**Min & Max % “value” per DC (When single DC has value, it’s counted in both the stats)**

****

**Min & Max % “value” of the total values per DC**

****

**What were the challenges?**

1. The major challenge was to analyze/normalize the data without having the full context of the data or the situation.
2. Not having the information on what the “value” column indicates was challenging. Also, there was no pattern to the values.

**What information you hoped would be there, but wasn't**

1. Information about some “Context” to the data.
2. Having Units of ‘Time’ column- milliseconds, seconds, minutes or information if it is a “timestamp” and if yes then the format it is in.
3. Information on what “Value” column represents.
4. Information on desirable outcomes. For example: Min or Max value, Number of Values per DC.