**What is a hard drive?**

a high-capacity, self-contained storage device containing a read-write mechanism plus one or more hard disks, inside a sealed unit

**What are the 5 key SMART stats for identifying Hard Drive failures?**

SMART 5 – Reallocated\_Sector\_Count.

SMART 187 – Reported\_Uncorrectable\_Errors.

SMART 188 – Command\_Timeout.

SMART 197 – Current\_Pending\_Sector\_Count.

SMART 198 – Offline\_Uncorrectable.

**How should we aggregate stats for those features?**

All features are essentially counters for different kinds of errors, taking the max for these values should be sufficient.

**What year should we begin our analysis?**

Data seems to be robustly filled from 2016 forward. It is more sparse beginning in 2015. We should use 01-01-2016

**What should be our cut off year for early failures?**

Reading for industry standards the expected life of a hard drive is 5 years.

based on BlackBlaze statistic failure rates dramatically increase after 3 years.