

|  |
| --- |
| On Hard Drives & Failure  KE 5107 CA 1 Report |
|  |
| Authored by:  Raj Baharat  Edwin Tam  Vignesh |

Contents

1. What in Backblazes is this about?
2. Understanding & Exploring Hard Drives
3. Preparing the Data
4. Models for prediction
5. Evaluations

# What in Backblazes is this about?

|  |
| --- |
| Context & Business Backblaze is a cloud storage provider for developers and IT departments. They run and manage data centers to make sure there’s enough capacity, redundancy, and functional hard drives to read, write, and store data.  As such, it is important that they know the “goodness” of each hard drive – so that they can replace faulty drives early and efficiently.  To achieve that, Backblaze uses SMART statistics of hard drives.  **What are SMART statistics?**  Also known as Self-Monitoring, Analysis and Reporting Technology, SMART statistics is a monitoring system included in hard drives that reports on various attributes of the state of a given drive.  There are about 70 SMART stats ([source](https://www.backblaze.com/blog-smart-stats-2014-8.html)) available… some of which can be used to predict the failure rate of a hard drive.  However, hard drives – even those from the same manufacturer – provide inconsistent SMART stats.  What’s the Business Problem?  As said earlier, Backblaze’s business is based on the “goodness” of their hard drives. They need to know when a hard drive will fail early and efficiently.  Thus, we can ask:   1. Which SMART stat contributes to failure or success rates of a hard drive? 2. What would be a good model to predict hard drive failure?   Obviously, we can use the answers to focus on a few SMART stats and models. With these questions in mind, we now look at the available datasets. |

Where do we get this data?

Backblaze logs Hard Drive Test Datasets on their [webpage](https://www.backblaze.com/b2/hard-drive-test-data.html). Each zipped file contains daily snapshots for each of their 40,000 Hard Drives. This also includes hard drive statuses (Failure or OK).

We have selected to use 2018 dataset instead of across the years. Of note, the 2018 dataset contains additional SMART attributes which we will not use in this CA.

# Understanding & Exploring Hard Drives

|  |
| --- |
| Subtitle Text Here To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |
| *“Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment”* |
| To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |

# Preparing the Data

|  |
| --- |
| Subtitle Text Here To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |
| *“Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment”* |
| To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |

# Building Models to Predict Good Hard Drives

|  |
| --- |
| Subtitle Text Here To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |
| *“Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment”* |
| To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |

# Did It Work (Or Evaluating our Models)?

|  |
| --- |
| Subtitle Text Here To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |
| *“Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment”* |
| To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |

# What in Black Blazes is this about?

|  |
| --- |
| Subtitle Text Here To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |
| *“Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment”* |
| To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.  Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need. |