Stephanie Nelson

March 28th, 2023

Professor Bemley

**Title**: “Statistics is the Grammar of Data Science—Part 1”, “Basic Components of a data set” and “Jupyter Notebook for Beginners: A Tutorial”

**Author**: Semi Koen and Benjamin Pryke

I. **Introduction**

The three articles Statistics is the Grammar of Data Science—Part 1, Basic Components of a data set and Jupyter Notebook for Beginners: Part 1 Tutorial helped reinforce information. Statistics is the Grammar of Data Science—Part 1 by Ms. Keon looks at statistics and foreshadows the ways we will see statistics in data science. Basic Components of a data set looks at data and explains how to break it down to then use it in the next stages of the data science pipeline. Jupyter Notebook for Beginners: Part 1 allow us to install and use Juypter Notebook in the early stages which is a credible application for data scientists.

II. **Basic Components of Dataset**

This article dives into what to look for when looking at a dataset. The use of visuals and text allow the audience to grasp understanding of the topics. Important concepts discusses was the difference between quantitative and qualitative data, acronyms and elements. The difference betwee qualitative and quantitative data is quantitative looks at how many while qualitative looks at how to identify an element. Acronyms are important in data science to save time and make assignments more productive. Elements and variables are what you look at in the data set. Thos article helped understand the specific name of items as well as the use behind them

III. **Statistics is the Grammar of Data Science**

Keon begins this article by explaining why a proficient understand of math is needed to become a data scientists. Keon then goes over different methods and ideas from statistics that will beneed to analyze data. Statistics is used in a various number of ways in data science. Keon breaks down data types into numerical, categorical and ordinal. Numerical is data represented by numbers, integers or decimals. Categorical is data represented by categories such as school subjects. Ordinal is data represented by ordered units. The use of mathematics in data science from data types to measures of central tendency. This was a refresher on math topics that will be needed to go through the data science pipeline process.

**III. Jupyter Notebook for Beginners: A Tutorial**

Pryke introduces us to Juypter Notebook as well as explains why it is so popular and useful in the data science industry. Jupyter Notebook first publishes back in 2010 and is able to function with code, visuzalization, data sets, mathematics and more. The article takes us through the installation of the application, creating out first notebook, keyboard shortcuts and other useful information. Some keyboard shortcuts mentioned are z to clear a cell, a to insert a cell above the current one you are in, b to insert a cell below the current one you are in and more. This is all useful information o read prior to working with Juypter Notebook.

V. **Conclusion**

In all, in this article, these articles provide readers with a great start to think like data scientists. We are reinforcing skills needed to tackle problems efficiently and learn shortcuts with applications/techniques needed in the profession. The articles are detailed as well as provide in depth understanding of the topics. Keon brushes up on needed statistic topics while Pyrke gives us a run through of Juypter notebook with background information.

Vi. **References**

Basic components of a data set.

<https://bsuonline.blackboard.com/bbcswebdav/pid-17985974-dt-content-rid-28398460_1/xid-28398460_1>

Koen, Semi (2019). Statistics is the Grammar of Data Science—Part 1, <https://towardsdatascience.com/statistics-is-the-grammar-of-data-science-part-1-c306cd02e4db>

Pryke, Benjaminn(2020). Jupyter Notebook for Beginners: A Tutorial, <https://www.dataquest.io/blog/jupyter-notebook-tutorial/>