LESSON HANDOUT

The Data Analytics Environment

How do we do it?

The data analysis process involves 4 stages.

Although each stage is crucial to proper analysis, they are not linear but interchangeable throughout the process.

- The **DEFINE** stage is where we find our questions. What are we looking for from this data?
- The **TRANSFORM** stage is where you prepare your data for analysis and consists of formatting and validating your data.
- In the **ANALYSE** stage, we look to discover answers from our data. We attempt to extract insights, notice patterns and relationships within our data.
- In the **COMMUNICATE** we tie our findings to the original query and presents a solution.

The context:

Data is everywhere! In our day and age, every process, system, media, and machine produce data. Your text messages, purchases, social media posts, and even the videos in this course are all data.

Our society's decisions, improvements, and processes are constantly driven by data. Targeted personalized advertisements, voice

recognition, social networking, and e-commerce are all enabled by data.

How did we get here?

4 factors that have contributed to the unprecedented availability and opportunity that is presented by the data environment today.

Data Collection: The growth of connected electronic devices in the last few decades has exponentially increased the amount of data we collect. Internet searches collect data daily. To put this into perspective, there are approximately 5.6 billion searches performed per day and that's on Google alone!

Data Storage: This data collection has been enabled by the amount of inexpensive and electronic data storage options. In 1956 IBM's first hard drive could store the equivalent of 1MP3 song. Whereas today, a desktop drive can store up to 760,000 MP3s at a time.

Processing Power: The rise in data has been accompanied by an exponential growth in the computing power required to process this data. To put this into context, we now have more computing power in our current Iphones than did the computer aboard the Apollo 11 when it landed on the moon in 1969.

Statistical and Programming methods: Many of the algorithms and statistical methods were invented by academics centuries ago. What has changed is technology and the ease and speed at which we can apply these methods today.

All these factors combined have produced an environment where every industry is becoming reliant on data and the pace of change just keeps

growing. As **data analysts**, we must be willing to be **continuously learning and ready to adapt** to this everchanging environment.