**CASE STUDY 1**

As I worked on the Cyclistic bike-share case study over the past two days, I have encountered

my fair share of excitement and challenges. My role, in short, is to focus on analyzing how

annual members and casual riders utilize Cyclistic bikes differently. The results of my analysis

will help guide the marketing strategy of how Cyclistic promotes its membership to the casual

riders. If done correctly, the director of marketing thinks it can have a huge return on investment

and accelerate future success for the organization. But it’s far easier said than done. There have

been some technical hurdles to overcome. For example, after first ingesting the historical bike

trip data, I have already discovered that there’s some concerns about the data’s quality. It’s a

large dataset, with many variables and potential missing values, outliers, and other

inconsistencies with how people use the dataset. So far, I’ve been busy coding and organizing the

data just for the purpose of cleaning – to make sure the end understanding reflects the reality of

what people use bikes for. Secondly, looking ahead, I’m curious how I will organize the same in

a way that’s meaningful: how will I tell the story by visualizing differences from the same,

separate group of casual riders and annual members? I understand the core findings and how

they are beneficial to the business of making it easier for people to ride bikes. But having my

findings communicate as clearly as possible is a challenge I will need to figure out before

presenting to stakeholders who are less data-intensive in their thought process.

Two technical questions present: 1. What are some of your best practices for cleaning and

preparing data for analysis, especially large datasets?

2. Are there specific metrics or KPIs I should focus on to demonstrate annual membership value

for a Chicago mental health organization?

­­