Good evening Sprocket Central,

My name is Mario Paul and I am a Data Science intern for KPMG.

I was tasked with assessing the three datasets on customer demographics, addresses, and transaction data in the past three months that we have received from you. This is a process that is done before we begin to work with any data we receive to ensure we can get the most accurate results and answers we are seeking.

As I was going through the datasets we received from you, I found a few things that I’d like to bring to your attention that could harm the results. Here are my notes that i’ve made.

**Customer Demographic:**

**Not Accurate:**

Jephthah Bachmann DOB claims he was born in 1843

**Not Consistent:**

Gender:

1 entry is ‘Femal’

1 entry is ‘M’

1 entry is F

**Not Valid:**

Gender:

88 entries are ‘U’

**Completeness:** The following Columns have missing / null values:

**Last\_name 125 (0.03125%)**

**DOB 87 (0.02175%)**

**Job\_title 506 (0.12650%)**

**Job\_industry\_category 656 (0.16400%)**

**default 302 (0.07550%)**

**tenure 87 (0.02175%)**

**Not Current:**

Two of the customers are deceased, Josey St. Quentin and Kurtis Morson.

**Relevancy:**

“Default column is not relevant.

**Customer Addresses**

**Not Consistent:**

Customer\_id’s not consistent

Address house / building numbers are not consistent. Some single digit numbers have a 0 in the front of them and some do not.

Some addresses are not correct. Some addresses start with the number 0

**Relevancy:**

Property\_valuation column is not revellant. Has nothing to do with address information.

Country column is not needed since all entries are from the same country.

**Transaction data**

**Completeness:** The following Columns have missing / null values:

**online\_order 360 (0.01800%)**

**brand 197 (0.00985%)**

**product\_line 197 (0.00985%)**

**product\_class 197 (0.00985%)**

**product\_size 197 (0.00985%)**

**standard\_cost 197 (0.00985%)**

**product\_first\_sold\_date 197 (0.00985%)**

**Not Current:**

Values are supposed to be transactions from the last 3 months. They are dated from 2017.

**Relevancy:**

There are 179 cancelled transactions where are not needed

After fixing all of these errors, the data will be ready for our team to begin the next step.

Thank you,

-Mario Paul