UNIVERSIDAD CENFOTEC

Data Analytics / Big Data 2018c-3

Program 5 Task 2

Data Science with Python

Prepare and Explore the Data

PRESENTA:

Alberto Madrigal Jiménez

ASESOR:

Gustavo Rojas

San José, CRC.   2020

Did you learn anything of potential business value from this analysis?

The answer is yes; when we applied the EDA to the dataset we could understand some important features in the data. As you can see in the details, the behaviors in between 2 or 3 variables show new data and outcomes to take better decision, although we were analyzing in the previous reports some fields, formats and others stuff with the framework, the way to show up the data with graphs, matrixes, and summaries improve the capacity to understand and make better decisions.

In the business perspective, I learned a lot to be open all the time to new results mainly, when is just one person in charge of the process, as is this case, the decision to take the best framework, the ETL process, and the analysis role in just one owner, probably improve the skills when the next step comes to be done, but, it requires that you need to come back to check again, fix something and then run the process several times.

What are the main lessons you've learned from this experience?

Without a doubt the best lesson I have learned with this specific case is to use a process of the continuous improvement, does not close the door of the opportunity to improve when new things are found. Keep moving forward, and don't falling in a buckle, I believe that, if you understand, you can do, and if you can do with understanding, you can find new ways to read the results and achieve better results.

What recommendations would you give to the Guido regarding your findings?

Based on the outcomes that we captured the recommendations to give to Guido are:

- Don't remove fields from the data; we can see as all the fields can improve when we apply the models.
- Don't remove the negative values; it has a significant mean when in the financial side it depends.
- Please use the new cluster in the data, because it help and improve the data model and the understanding over the data, mainly, when there are some values that they mean the same value, for example when 4,5,6 and 0 are other.
- Move forward to the next step (apply the model), and create a second ETA to analyze the outcomes, it can help to make the best decision as we did it in this analysis.