|  |  |
| --- | --- |
| Name Of The Student | Vaishnavi G |
| Internship Project Topic | Build a Classification Model for Drug Trials Dataset |
| Name of the Organization | TCS iON |
| Name of the Industry Mentor | Himdweep Walia |
| Name of the Institute | SRM Institute of Science and Technology |

|  |  |  |
| --- | --- | --- |
| Date | Day # | Hours Spent |
| 25/10/2022 | 14 | 5 hours |
| Activities done during the day:  Learned about data transformation process using a live Python use case scenario  Five steps that are mandatory in any research with data:   1. Set your analytics goal – what is a real-life problem that needs research 2. Collect data – find sources of data 3. Prepare data – check data quality, integrity, completeness and perform data cleaning (avoid GIGO – “Garbage In, Garbage Out” situation) 4. Analyze data – the fun part 5. Interpret results – understood results, do a reality check, go back to phase 2 or 3 if needed, do data visualization, prepare the next actions, etc.   **Data transformation using Python:**  **Step 1 :** Create Your Python Script  IMG_256  **Step 2 :** Import a Pandas Module    To load the CSV files into your notebook, you need to import a Pandas module, which is an add-on file that extends Python’s basic functionalities. To do this, simply type “import pandas as pd”.  IMG_256  **Step 3:** Load file content  pd.read\_csv( ‘/home/jacekpolewski/tutorial\_files/data\_file\_1.csv’, sep=’;’)  **Step 4:** Assign Data Content to a Variable  IMG_256  **Step 5:** Inspect Data Using the Head() Function  After loading your data and assigning it to a variable, it’s a good idea to inspect the data to see if all of the content has loaded correctly. You can do this by looking at the first few rows of the data using the Python head() function.  **Step 6:** Consolidate multiple Datasets  IMG_256  **Step 7 :** Data Transformation/Preparation  If you need to modify your data or transform it into your desired format before extracting valuable insights, Python and Pandas make it as fast and easy as possible to do so.  If you need to create a new column, just type in and execute data\_comb[‘Test’] = ‘hello’. When you now inspect the data, you will notice a new column called ‘Test’ filled with values ‘hello’.  IMG_256  Reference:  #1 - <https://www.tibco.com/reference-center/what-is-data-transformation>  #2 - <https://towardsdatascience.com/data-transformation-and-feature-engineering-e3c7dfbb4899> | | |