*Project Team 1.-*

**Procurement database with Requisition and Purchasing Order information**

***Extract:***

Original files consist of 4 files in csv format coming directly from the eProcurement tool:

* Approval data
* Exchange rates
* Order header
* Order Line
* STAR data

***Transform:***

Each of the files required to be cleaned up on their column headers, eliminate some columns and rearrange their order to give them a better sense. The team had difficult time on cleaning data on 2 tables:

* Approval data: the main constraint was to convert the dates-time columns to a date-24h format and then to have Postgres accepting the values of this conversion. At the end the team decided to perform a second conversion of the date-24h to string to help Postgres to accept the information.
* Order Lines: the main constraint was to delete the values in this table that were not appearing in the order\_header table so we could link these columns through Foreign Keys, what was decided was to perform INNER JOIN in Postgres then export the result of this inner join to a csv document that then will be read using pandas

***Load:***

Loading was done through an engine directly to Postgres, team decided to create an additional table called old\_order\_lines so we can keep track of records that don’t have the right data quality and investigate further the main reason.

***Team members:***

Andrés GUTIERREZ

Lina DAVILA

Paulo YOUNG

Purvi PATEL