|  |  |
| --- | --- |
| ETL Project | Summary  We decided on Books for our Project.  Created by  Alejandro Gomez  Brook Forrester  Jessica Dean.  ETL Project |

For our project we decided on books. We used Kaggle as the source of data. We wanted to do Food Trucks as our data but sadly there were no working API keys. So, we switched over to books as all of us on the team liked books and thought that it would be a fun data frame to work with. We had a few issues with getting the data to load into PostG. And mostly stuck into Pandas where it would let us join and merge, while on PostG it kept over writing the data. We normalized the data into three tables. We listed the Tables as Genre, Price, Reviews\_and\_Ratings as they were the most important part of the data frame and gave the best results.

Our questions and answers are below.

Our kaggle link is <https://www.kaggle.com/sootersaalu/amazon-top-50-bestselling-books-2009-2019>

We normalized the data to enter it into the PostG database called best\_book\_db.

We also made it a relational database.

We read the Jupyter notebook and the CSV and normalized the data into three tables then we created the table in PostG SQL we filled the PostG tables with the data using SQL Alchemy. Then we verified the data in PostG using Pandas and within PostG.

Photos are below.





