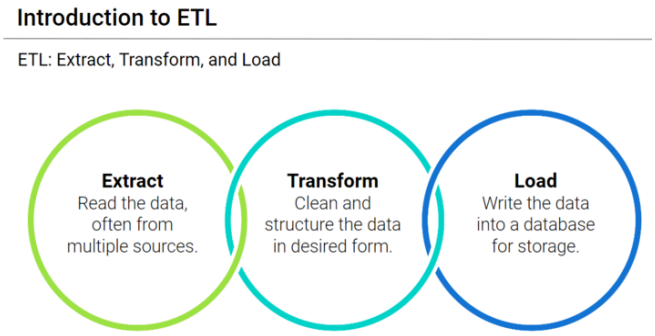
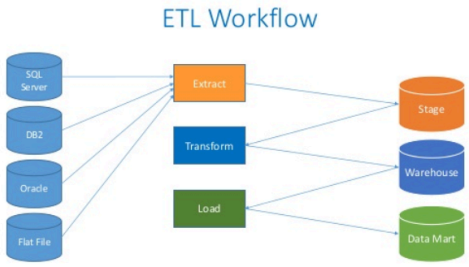




ETL stands for extract, transform, load. It is the process of manipulation and retrieval through different databases, transforming it to ensure consistency and accuracy, then loading it into a destination database for analysis and use. Data can be extracted from a variety of sources such as APIs, files, and a variety of data structures. This process is crucial for making data readable and actionable.

For this project, the dataset that we are working with is related to crowdfunding campaigns. using regular expressions to find patterns and extract data from text and string data.

First, we extracted the data from an excel file, making the data into a raw but accessible format. Next, we transform the data by coding to manipulate the dataframes, making it clean and in the right format. After that, we loaded the data into our sql database. Finally, we create and run our queries to generate the insights and more about our data. This process helps us to better ensure our data is ready for analysis and well-organized.

For the analysis our group asked several questions of the data, such as:
- Is there a correlation between the amount funded and its success rate?
- How many backers_count had a value of greater than or equal to 300?
- Does the region affect the amount of investment each project received?

To visualize these insights, we created a few charts and graphs:

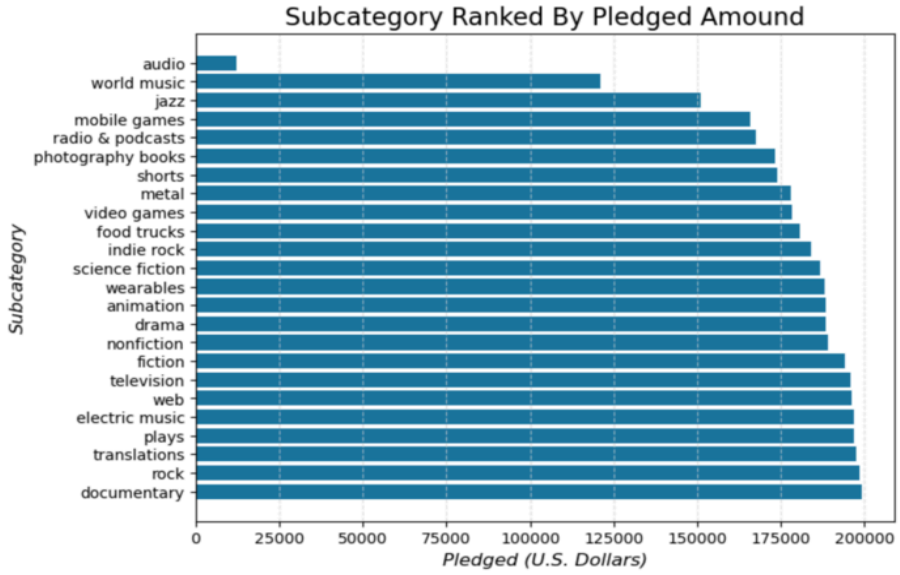
- A scatter plot to show the backers count of different crowdfunding projects grouped by cf_id.
- A pie chart to represent the highest funded sub-category.
- A line graph to illustrate the trends of successful projects and how does that correlate with the region.

- Biases and Limitations

One limitation is that we only have the average donation. Additionally, the data doesn't include recent dates, which might miss current trends which can lead us to overlook certain outliers and impact the overall analysis. Outliers can skew the mean, making it not a direct representative of the data spread.

- Conclusion: Overall, our analysis provided insights into how these factors correlate with project success and failure rates, giving us a deep dive into business intelligence, project trends and helping us work with big data sets, empowering us to be ready for real world events.

Here are some visualizations that we came up with:



-