Assignment No. 1

Task A

According to data located here please apply the following required actions.

- Feature engineering (at least select 8 variables + Churn variable)
- Convert total charges into categories (discretization)
- Missing values treatment
- Label encoding
- Standardization

And then get some information from the entire data such as:

- Relation between each variable and churn
- Relation between each variable and the others

Then please answer the following questions from the data:

1- How many male customers churned from our company?
2- What is the ratio between males and females in our company?
3- What is the ratio between churned and others in our company?
4- How many senior citizens churned from our company?
5- What is the id of the biggest in total charges?
6- What is the correlation between monthly charges and total charges?
7- What is the average payment per month of all users?

Assignment No. 1

Task B

According to data located here , please apply the following required actions.

The following points to be answered:

1. What are the genres that have the highest average rating?
2. Who are the highly rated directors?
3. Who are the highly rated actors?
4. Who are the most profitable directors?
5. Who are the most profitable actors?
6. Which countries have higher average rating?
7. Which countries produced more movies?

Dataset Specifications:

It is a dataset from the TMDB (The Movies Database) website for ~5000 movie titles separated into two files:

Movie Metadata

- Movie_id — the movie id
- Title — the movie title
- Original_language — the movie language
- release_date — the movie date
- budget — the movie budget
- revenue — the movie revenue
- runtime — the movie runtime in minutes
- vote_average — the movie TMDB average rating
- vote_count — the movie TMDB rating users count
- popularity — the movie TMDB popularity score
- genres — the movie genres separated by a pipe
- keywords — the movie keywords separated by a pipe
- production_companies — the movie companies separated by a pipe
- production_countries — the movie countries separated by a pipe

Movie Cast and Crew

- movie_id — the movie id
- director — the movie director name
- producer — the movie producer name
- actor_1 — the movie actor_1 name
- actor_2 — the movie actor_2 name
- actor_3 — the movie actor_3 name