Project Idea #1:

* **Topic (Title):** Art Expert
* **High level description of project:**

Recognizing the genre and style of a piece of art. Data will be collected using web scraping the WIKIART. The ideal dataset will include, for an art work, picture, genre, style, author, year, etc. The goal of the project is to through machine learning to on features of artworks: theme, color, tone, material, form, etc and to be able to classify an art work to its most possible art style.

**Data Source(s):** WIKIART

* *Optional* **Potential Future Employer**:

Project Idea #2:

* **Topic (Title):** European soccer data EDA
* **High level description of project:**

The dataset contains 25k+ matches, players & teams attributes for European Professional Football, detailed match events (goal types, possession, corner, cross, fouls, cards etc...) for +10,000 matches, betting odds from up to 10 providers. The dataset can be used to predict game results, player’s performance score, relationships between betting odds and game results, etc.

* **Data Source(s):** <https://data.world/data-society/european-soccer-data>

Project Idea #3:

* **Topic (Title):** Amazon Customer Reviews Dataset EDA
* **High level description of project:**

Conducting EDA on the Amazon Customer Reviews dataset. The goal is to explore the dataset, find out the relationship between customer reviews and market.

* **Data Source(s):** <https://registry.opendata.aws/amazon-reviews/>

*The dataset contains the customer review text with accompanying metadata, consisting of three major components:*

1. *A collection of reviews written in the Amazon.com marketplace and associated metadata from 1995 until 2015. This is intended to facilitate study into the properties (and the evolution) of customer reviews potentially including how people evaluate and express their experiences with respect to products at scale. (130M+ customer reviews)*
2. *A collection of reviews about products in multiple languages from different Amazon marketplaces, intended to facilitate analysis of customers’ perception of the same products and wider consumer preferences across languages and countries. (200K+ customer reviews in 5 countries)*
3. *A collection of reviews that have been identified as non-compliant with respect to Amazon policies. This is intended to provide a reference dataset for research on detecting promotional or biased reviews. (several thousand customer reviews). This part of the dataset is distributed separately and is available upon request.*