**Project Description**: There has been a lot of talk about how technology is our only

hope of breaking the cycle of poverty and creating opportunities and wealth for

underprivileged populations and marginalized groups. We want to use data to

create clusters and predict the change over the next 5 -10 years.

**Technologies:** Python, Pandas, Web Scraping, API, SQL, NoSQL, Big

Data, Machine Learning (Possible building an App with Machine Learning),

Tableau to Visualize results.

**GitHub link:** <https://github.com/vihaveman/Breaking_Barriers>

**Outcome:**

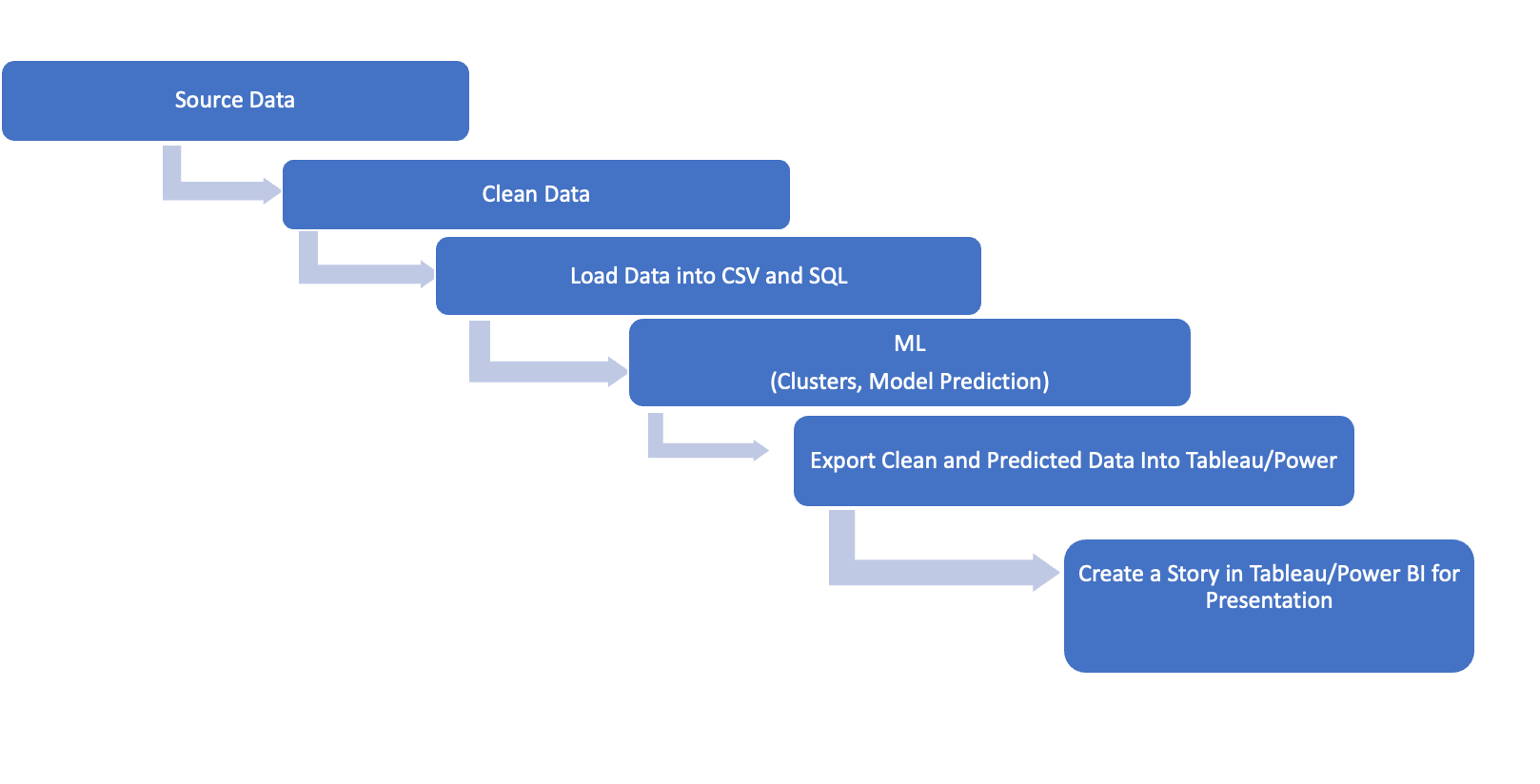
WE are the Data team leads for the Poverty Project and we would like to do the following:

* Source and Clean the data
* Load the Data into a SQL DB for quick look-up
* Perform Machine and Deep Learning
* Visualize the Data
* Present the findings

Have a model that predicts the economic change for these

target groups. Tableau Visualization showing this change.

**Expected WorkFlow**

****

**Timeline:**

| **Step** | **Description** | **Responsibility** | **Due date** | **Status** |
| --- | --- | --- | --- | --- |
| 1. Data Search | Dataset related to the technology level in countries. | Violetta | July 22, 2023 | Done |
| 1. Data ETL | Sixteen csv datasets with data from  Clean half Data Set  Merged data | Violetta  Zhibek  Violetta  Zhibek  Violetta | July 25, 2023 | In process |
| 1. Model development  * Clusters * predict the change | Delpoy Machine learning codes | Violetta  Zhibek | July 28, 2023 | Not started |
| 1. Data Visualisation | Create Tableau DashBoards and Story | Violetta  Zhibek | August 1, 2023 | Not started |
| Testing | Test | Violetta  Zhibek | August 2, 2023 | Not started |

**Data Search:**

Primary information source: The World Bank offers a vast array of economic and financial data concerning various countries worldwide. To access their comprehensive database, we visited the World Bank's official website at https://data.worldbank.org/.

**Data ETL:**

The data comprises numerous CSV datasets, each containing country-specific information on indicators related to the technological progress of those countries. Following the Extract, Transform, and Load (ETL) process, we obtained the resulting cleaned data as follows:

1. Automated teller machines (ATMs) (per 100), average for the period 2011-2021, float
2. GDP per capita growth (annual %), average for the period 2011-2021, float
3. High-technology exports (current US$), average for the period 2011-2021, integer
4. Communications,computer (% of service imports), average for the period 2011-2021, float
5. Fixed broadband subscriptions (per 100 people), latest for the period 2011-2021, float
6. Individuals using the Internet (% of population), latest for the period 2011-2021, float
7. Labor force total (people), average for the period 2011-2021, float
8. The later half: TBD

**Model development**

Our objective is to utilize unsupervised learning to cluster countries based on their technological advancement. The dataset (denoted as X) comprises all the variables we collected, cleaned, and merged. This approach will aid in gaining insights into the most significant features contributing to technological advancement among countries.

Next, we plan to employ supervised learning by using the poverty dataset as the target variable (y). By combining this dataset with the technological advancement data (X), we can create a model that predicts which countries are more likely to effectively tackle poverty based on their level of technological development. This approach aims to leverage the relationship between technological advancement and poverty alleviation to inform policy decisions and interventions for socio-economic progress.

**Data visualization**

To visualize the most interesting outcomes of the research, we can utilize powerful data visualization tools such as Tableau and/or Power BI.

Cluster Analysis: This could be a map or scatter plot, where countries with similar technological progress are grouped together and represented with distinct colors or markers.

Correlation Analysis: Visualize the correlation between different technological indicators and poverty rates using chareplot or correlation matrices.

Predictive Model Performance: We can use bar charts or line graphs to compare the model's predictions with actual poverty data, demonstrating its effectiveness in identifying countries with a higher likelihood of tackling poverty.

Regional Comparison: Compare different regions' technological advancements and their corresponding poverty rates using stacked bar charts or bubble maps.