**Author of project:** Jakov Vodanović

**Title of the project:** Analysis of Bank Transactions

**Database:** [Bank Customer Segmentation](https://www.kaggle.com/datasets/shivamb/bank-customer-segmentation)

**Description of the project:** Over one hundred and fifty thousand clients of an Indian bank have conducted over a million transactions, and details of these transactions have been stored in this dataset. It includes the client's date of birth, residence, gender, account balance, as well as the time and value of the transaction. The data was initially disorganized, and the first part of the project was basically data cleaning. Following data cleaning, I’ve analyzed the data to address the following questions:

* Sum and average of transactions over time:

1. How does the average transaction value and sum of transactions vary over time?
2. As time progresses, the data becomes fragmented. How can we fill in the "gaps," and what method can be employed for this purpose?
3. Is there a weekly pattern in transactions?

* Money laundering prevention:

1. On average, how many transactions does a client conduct?
2. Are there clients who perform excessively large transactions or an exceptionally high number of small transactions? What conclusions can be drawn about them?
3. What is the correlation between transaction size and account balance?

* Grouping clients based on transaction size:

1. Can we group them by gender? Is there a significant difference in the average transaction amount between genders?
2. Can we group them by age? Is there a significant difference in the average transaction amount among different generations?

* Analysis of bank activity by cities:

1. What is the number of transactions per city?
2. Can we deduce the bank's activity in cities solely from this number? If not, what is the alternative?

**Techniques/Algorithms:**

1. Exploratory Data Analysis (EDA)::
   1. Visualizations (line charts, histograms, pie charts, etc.);
   2. Descriptive statistics.
2. Statistical significance:
   1. Discussion of appropriate models;
   2. Hypothesis testing using ANOVA, t-test.
3. Time series analysis:
   1. Data prediction;
   2. Data decomposition.