|  |  |
| --- | --- |
| **DATE** | **31/10/2023** |
| **TEAM ID** | **687** |
| **PROJECT NAME** | **AI Driven Company Registration trend Prediction and Exploration** |
| **DEVELOPMENT** | **02** |

**Project Goal:**

The goal of this phase is to process the company registration data for time series forecasting, you'll typically need to perform several steps, including data loading, cleaning, and transformation.

**Dataset:**

In this document we guys are here to discuss the loading and pre-processing of the dataset to make a better prediction in the future.

**Source of Data:**

We got our dataset from the SKILL UP website for this project

**Exploratory Data Analysis (EDA):**

● Perform basic data exploration to understand the data's characteristics.

● Visualize the time series data to identify trends and patterns. Feature Engineering:

● Create additional features, such as lag features (past values) or rolling statistics.

● These features can provide more information for forecasting.

**Python Script:**

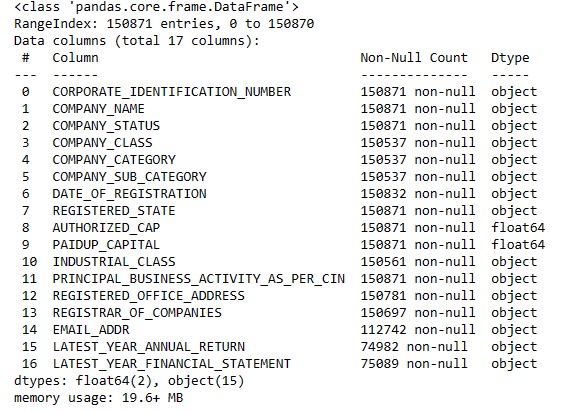
**Step-1**

The "import" statement in Python is used to bring external modules or libraries into your code, making their functions and classes available for use. It enables you to extend Python's functionality by accessing the features provided by these modules.

****

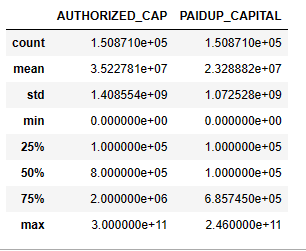
**Step-2**

df.info()` in Pandas provides a concise summary of a DataFrame's structure, including column data types, non-null counts, and memory usage, aiding in data exploration and quality assessment

****

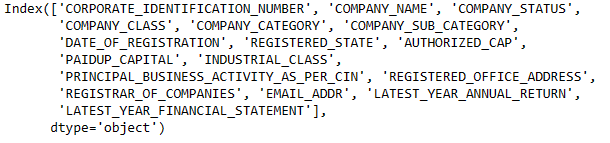
**Step-3**

df.describe()` in Pandas generates descriptive statistics for numeric columns in a DataFrame, including count, mean, standard deviation, minimum, and maximum values, providing a quick overview of the data's central tendencies and spread.

****

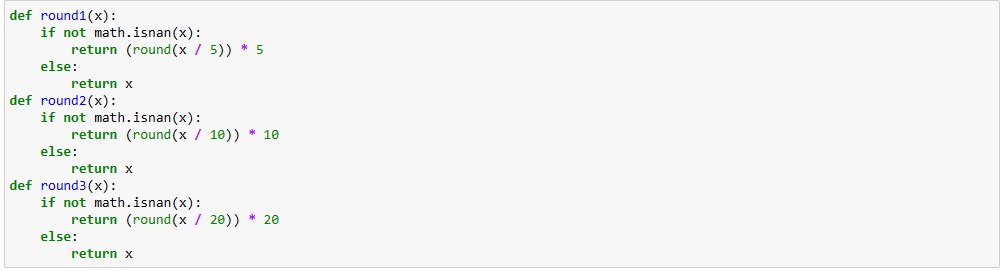
**Step-4**

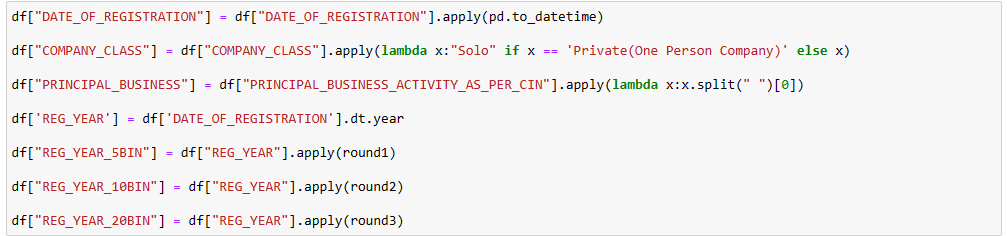
`df['column']` is used to access a specific column in a Pandas DataFrame, allowing you to retrieve, manipulate, or analyze the data within that column. Replace 'column' with the actual column name you want to work with.

****

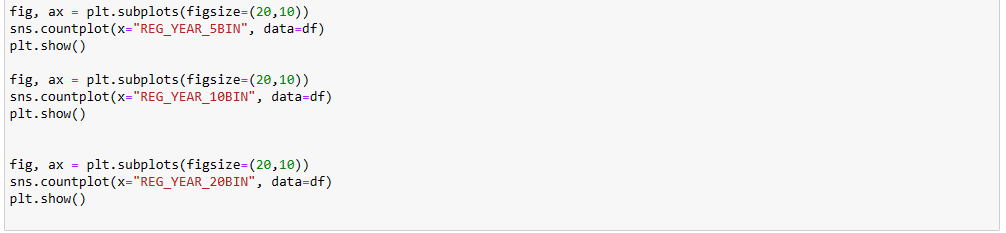
**Step-5**

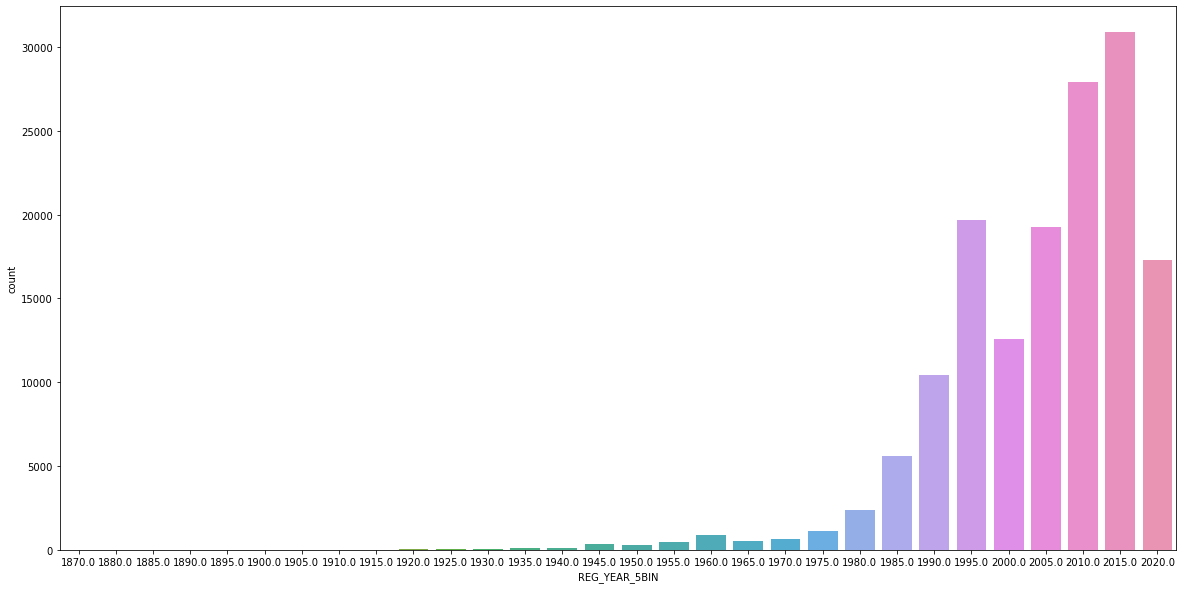
**Data pre-processing**

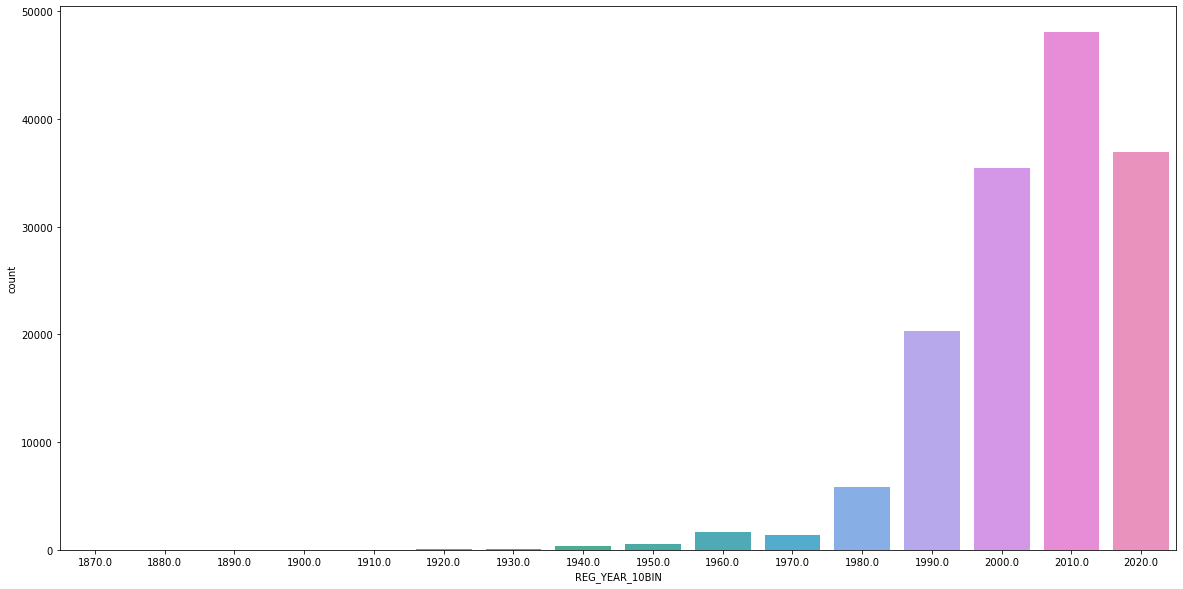
****

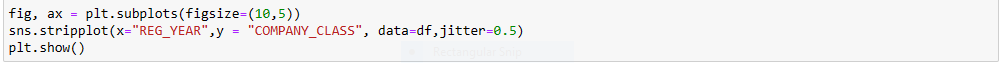
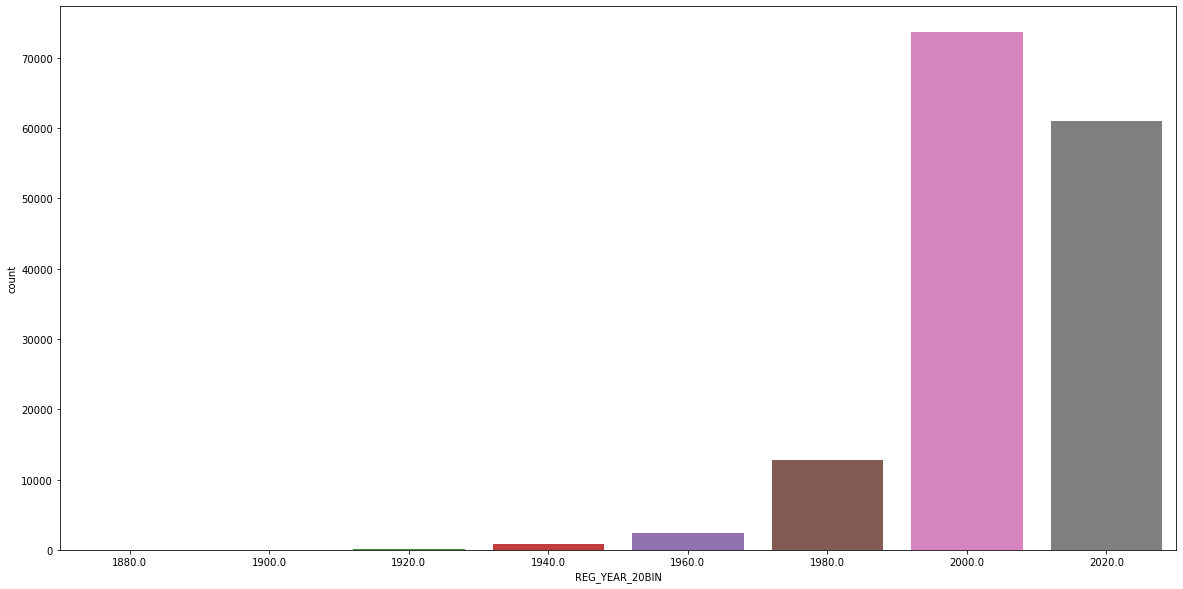
****

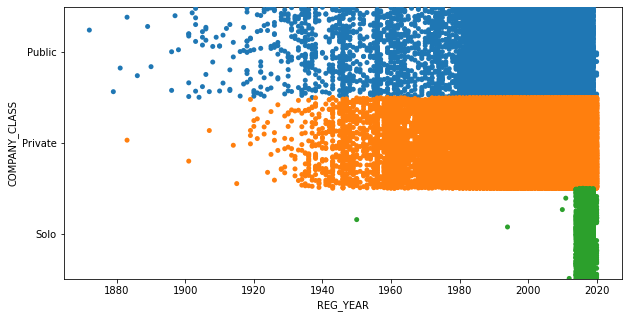
**Data visualisation**

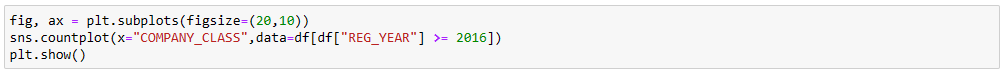
****

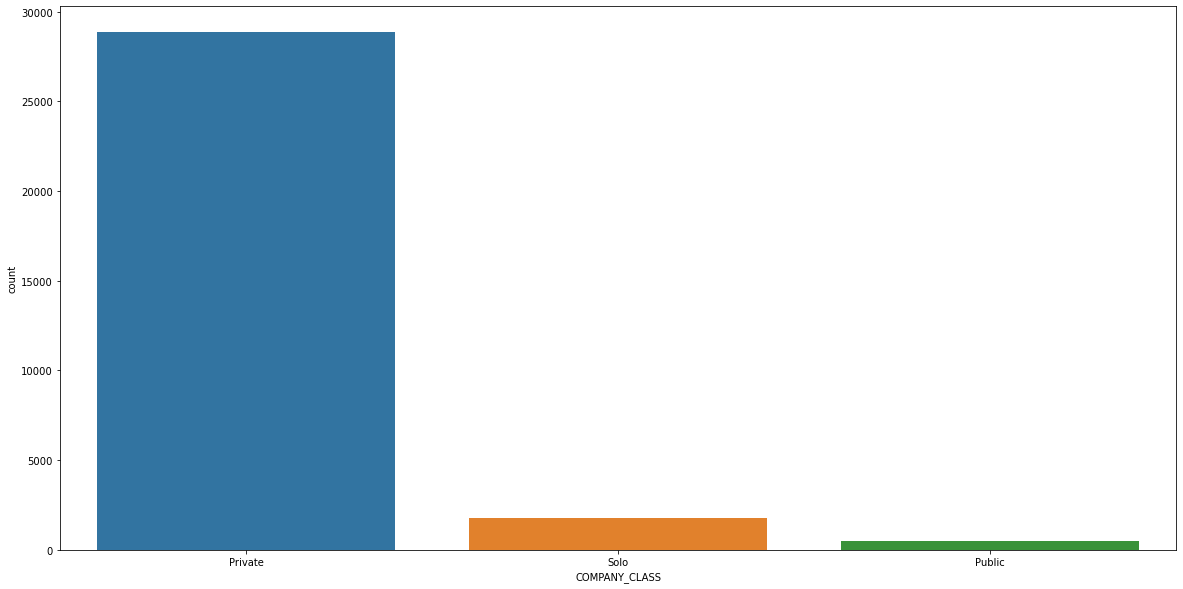
****

****

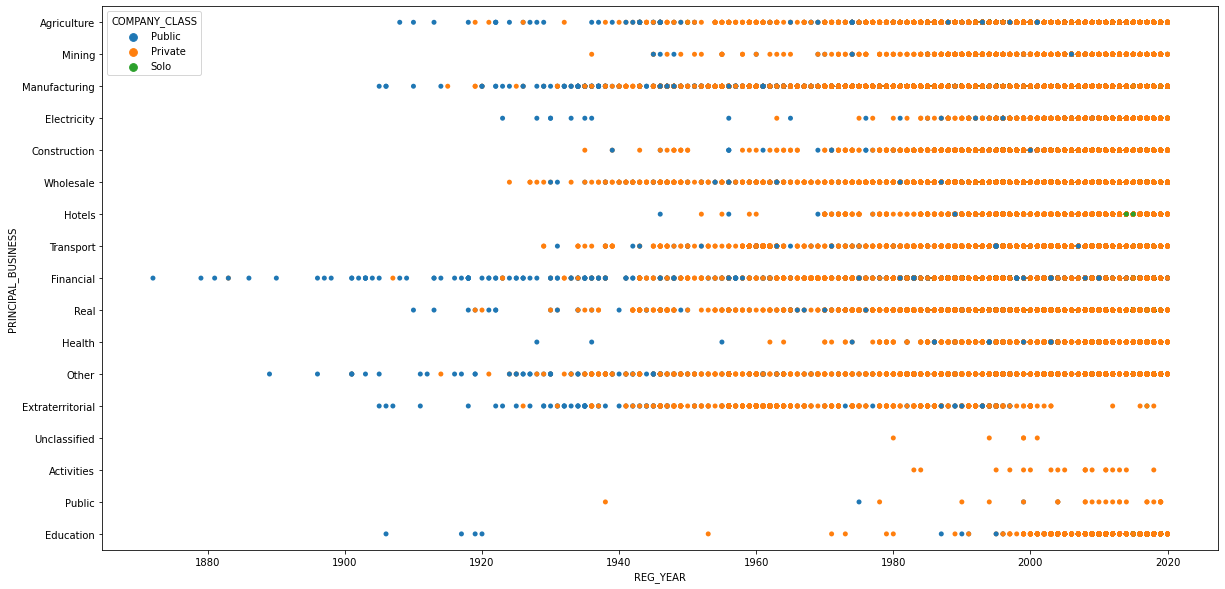
****

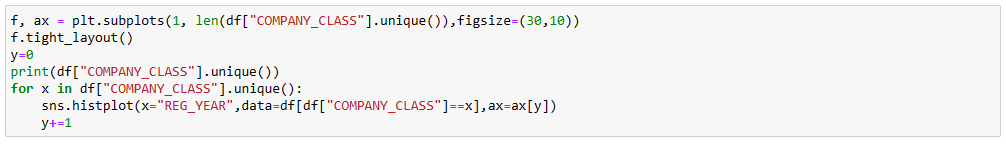
****

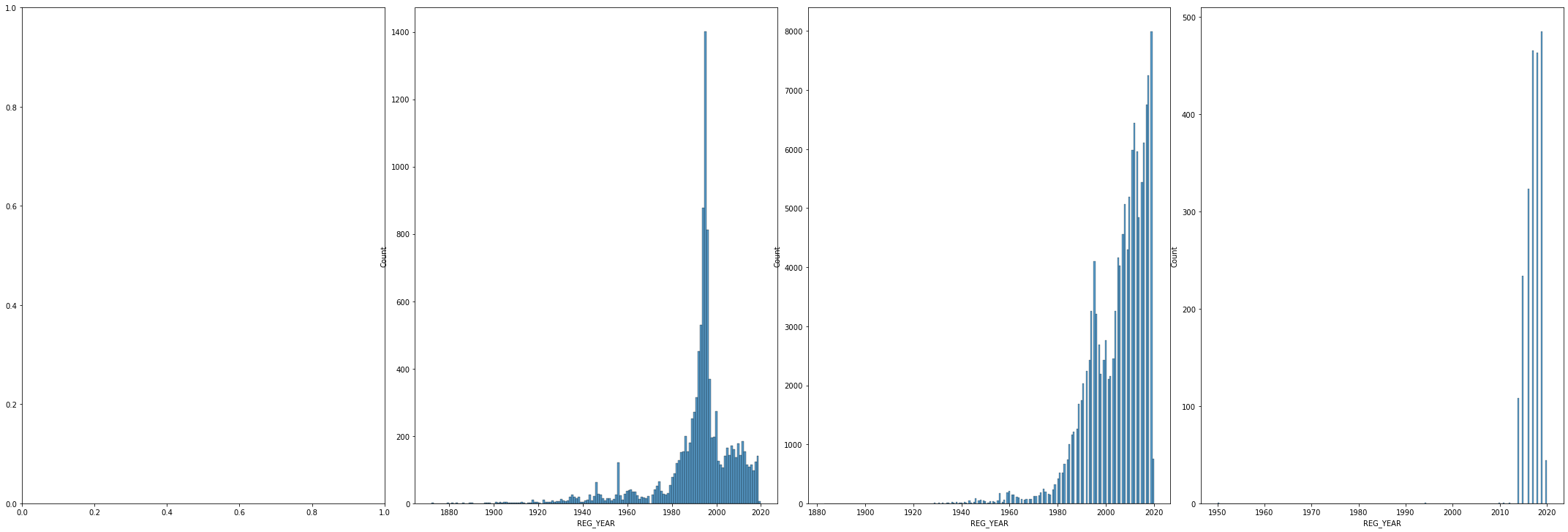
****

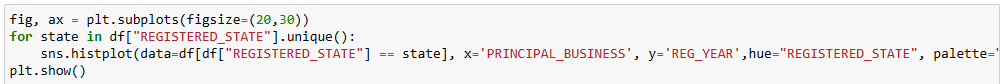
****

****

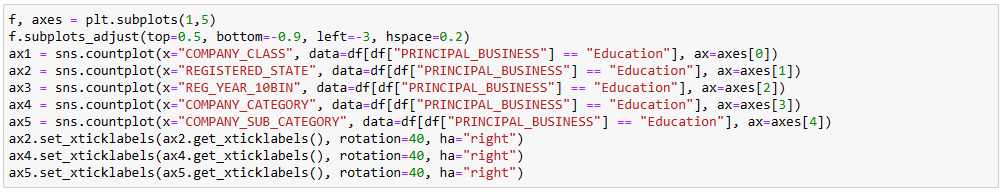
****

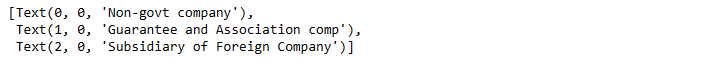
****

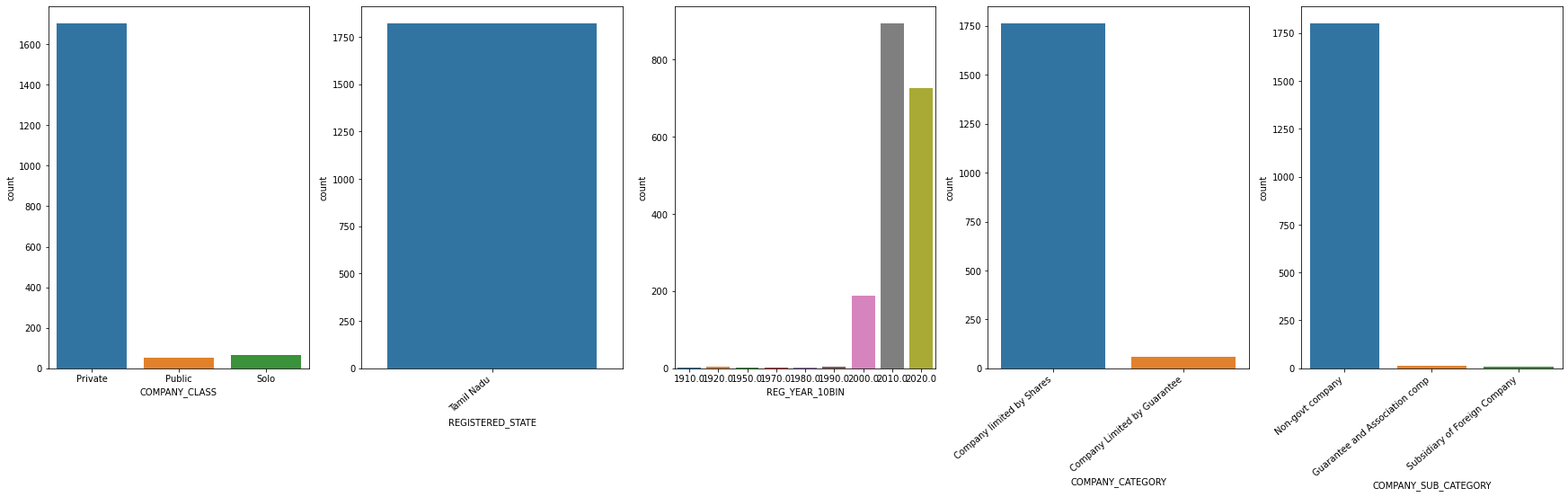
****

****

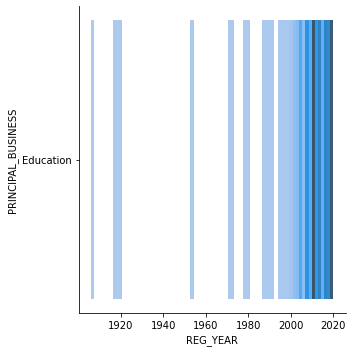
****

****

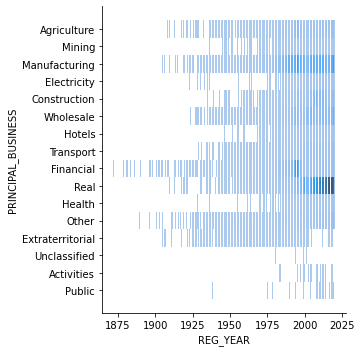
****

****

****

****

****

****