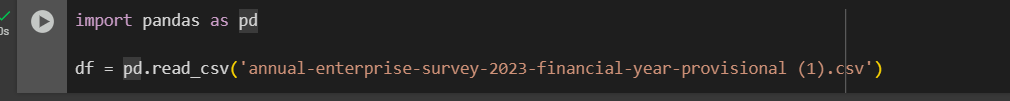
**Case study – Data Cleaning and Pandas**

**Annual enterprise survey: 2023 financial year (provisional) – CSV**

1. **Loading Data in Pandas DataFrame**

import pandas as pd

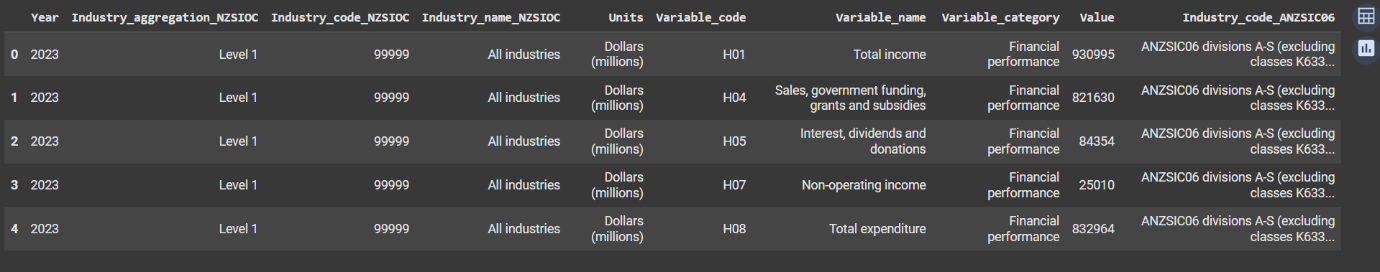
df = pd.read\_csv('annual-enterprise-survey-2023-financial-year-provisional (1).csv')



1. **Printing rows of the Data**

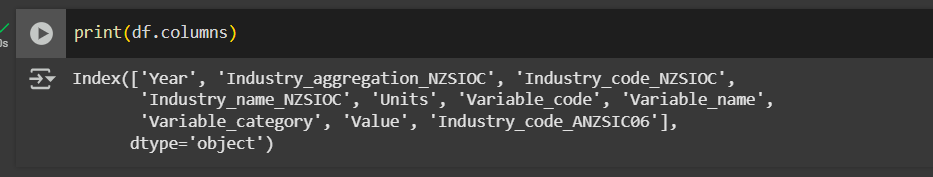
# Show first 5 rows

df.head()



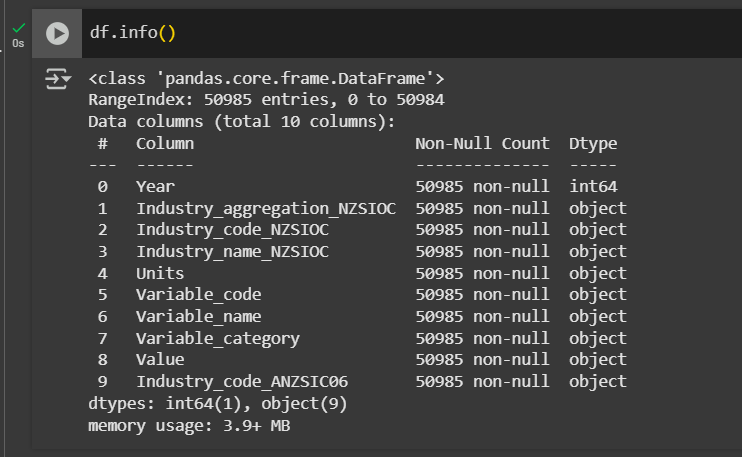
1. **Printing the column names of the DataFrame**

* print(df.columns)



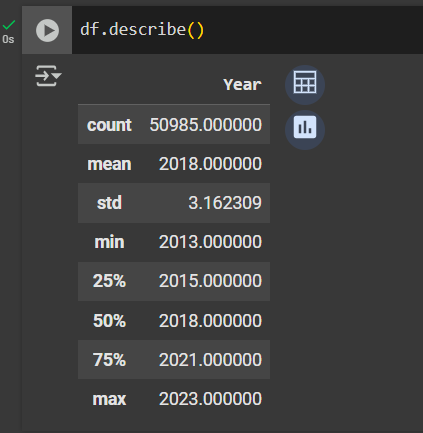
1. **Summary of Data Frame**

* df.info()



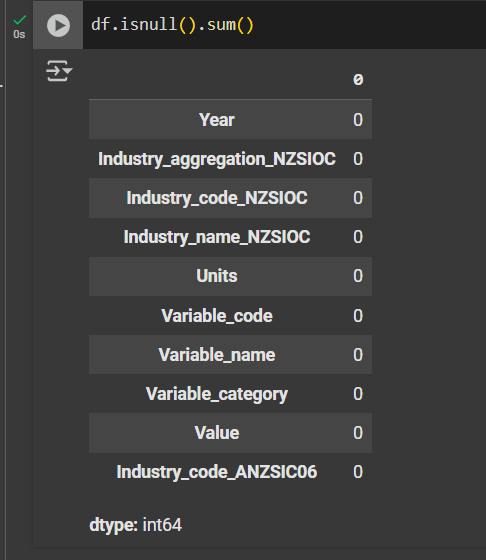
1. **Descriptive Statistical Measures of a DataFrame**

* df.describe()



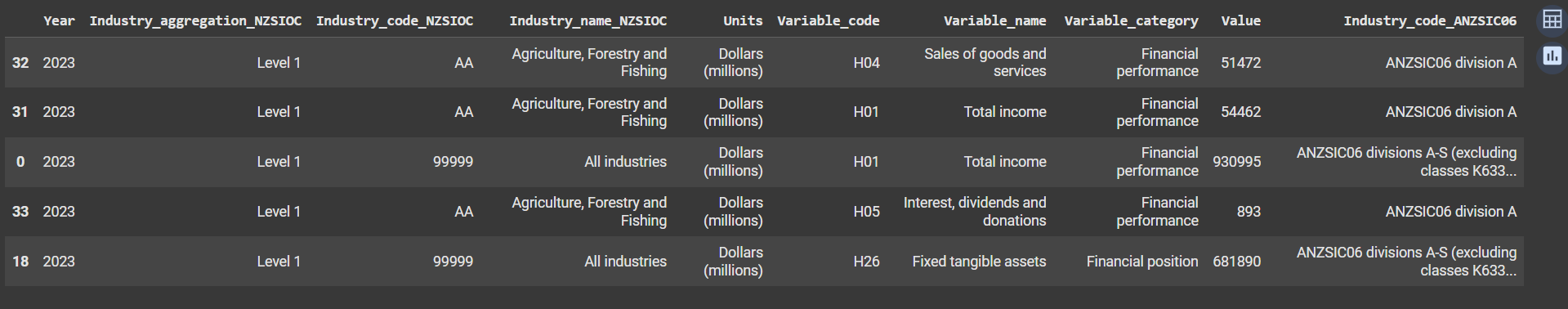
1. **Missing Data Handing**

* df.isnull( ).sum( )



1. **Sorting DataFrame values**

df.sort\_values(by='Year', ascending=False).head()



1. **Visualizing DataFrame**

import matplotlib.pyplot as plt

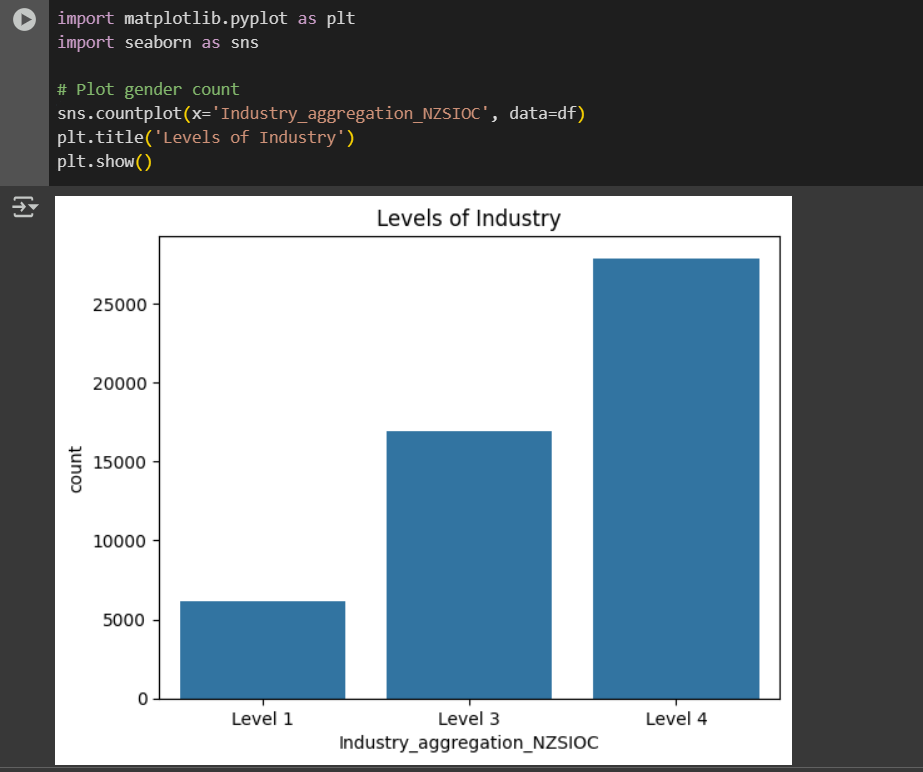
import seaborn as sns

# Plot gender count

sns.countplot(x='Industry\_aggregation\_NZSIOC', data=df)

plt.title('Levels of Industry')

plt.show()

****

1. **What is the number of columns in the dataset?**

num\_columns = df.shape[1]

print("Number of columns:", num\_columns)

1. **Print the name of all the columns.**

print("Column names:", df.columns.tolist())

1. **How is the dataset indexed?**

print("Index type:", type(df.index))

1. **What is the number of observations in the dataset?**

num\_observations = df.shape[0]

print("Number of observations:", num\_observations)

****

**Submitted by**

Siva Balan T