

مجتمع شاي للذكاء الاصطناعي

تقرير عن Task

الاسم : سارة عباس

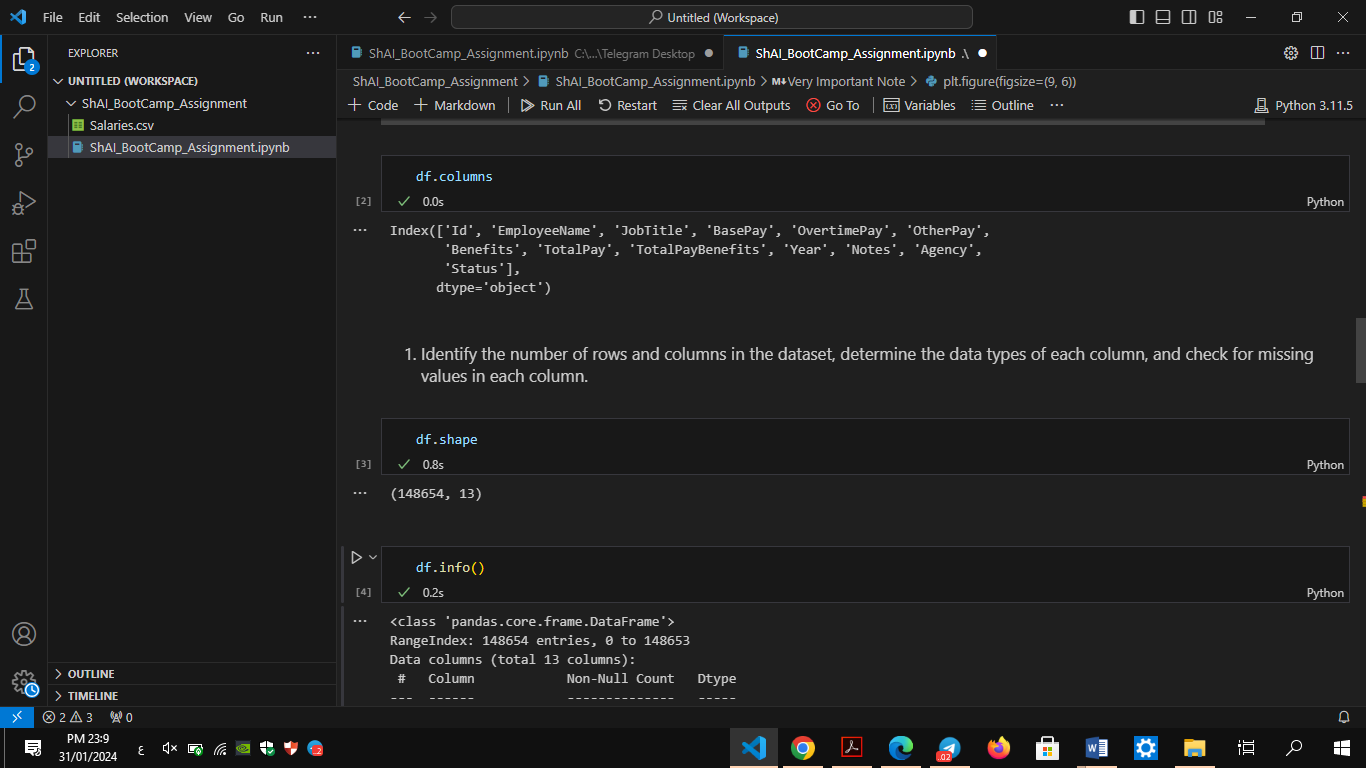
تم الإعداد بتاريخ 2024/1/22

1. **Identify the number of rows and columns in the dataset, determine the data types of each column, and check for missing values in each column:**

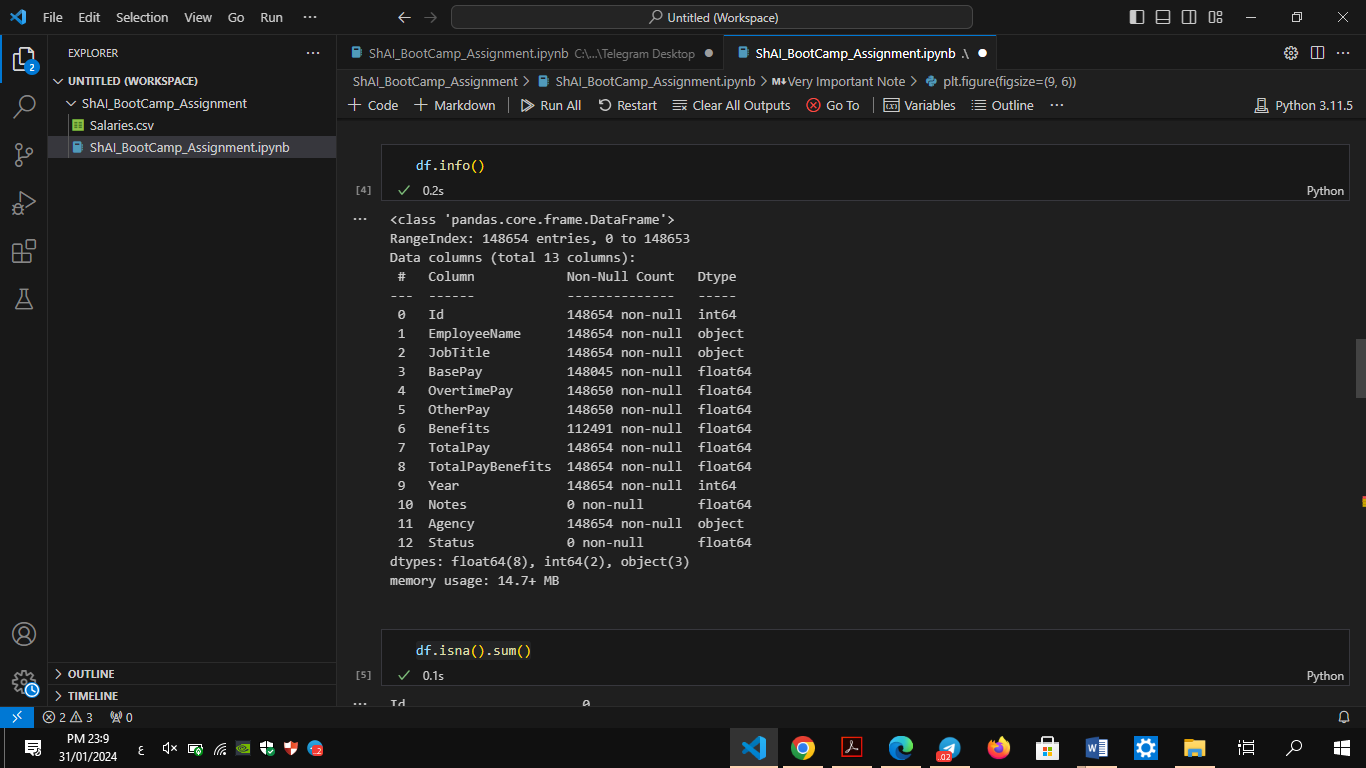
بعد قراءة مجموعة البيانات من ملف excel من خلال مكتبة pnadas

وتحويلها الى data frame نقوم بطباعتها

فنلاحظ ان data frame مكونة من 148654 سطر و 13 عمود



باستخدام التابع info() نقوم بتحديد نوع البيانات لكل عمود وحساب عدد الحقول الغير فارغة



ثم نقوم بالتحقق من القيم المفقودة في كل عمود من خلال التابع isna() فنلاحظ ان العمود BasePay يوجد فيه 609 قيمة مفقودة

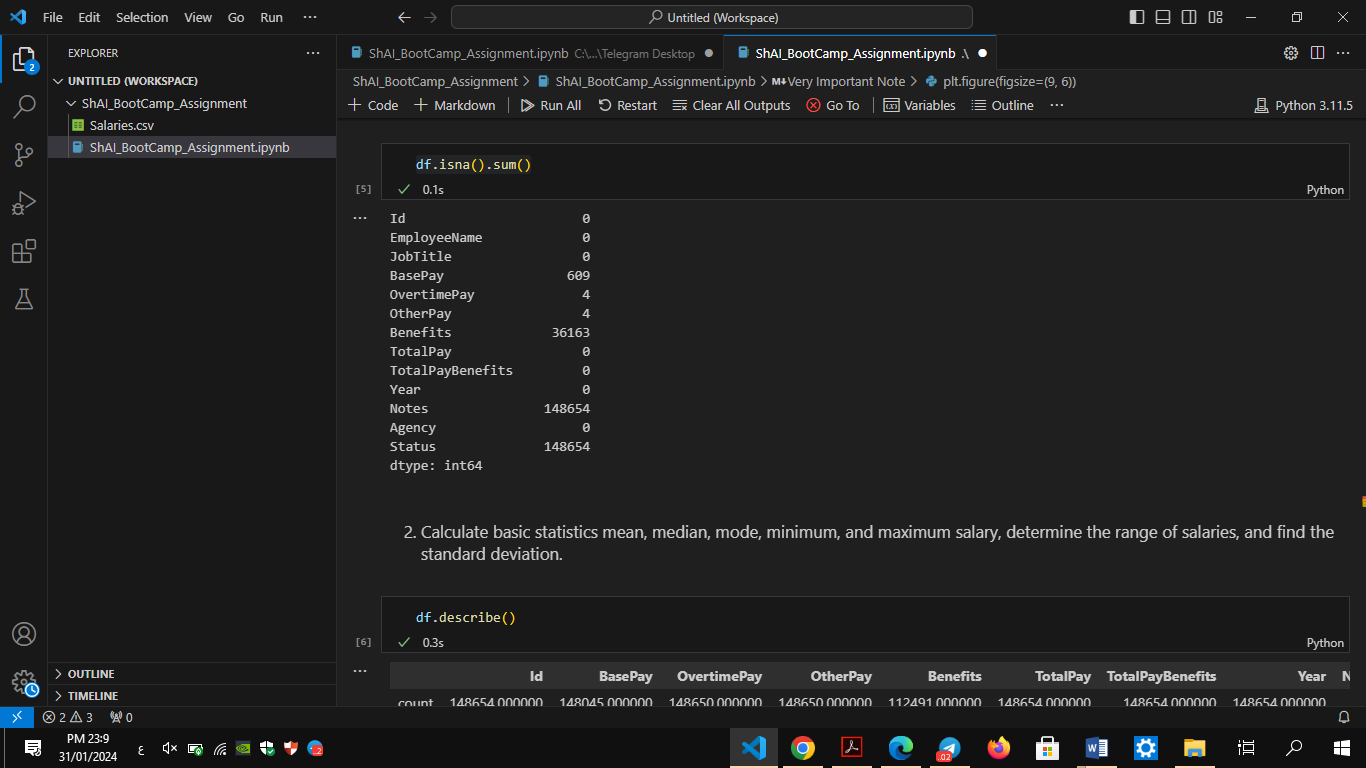
والعمود OvertimePay فيه 4 قيم مفقودة

و العمود OtherPay فيه 4 قيم مفقودة

والعمود BeneFits فيه 36163 قيمة مفقودة

وكذلك العمود Notes فيه 148654 قيمة مفقودة

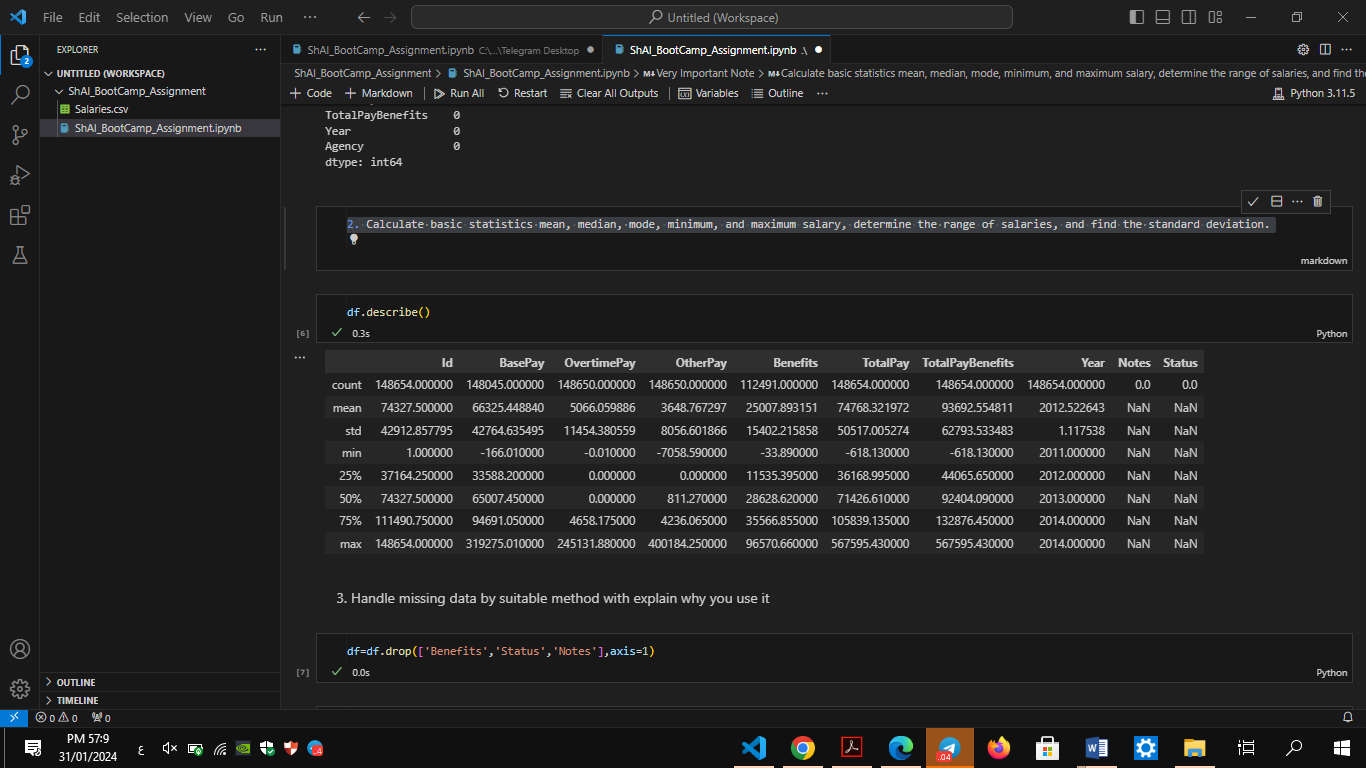
وأيضا العمود Status فيه 148654 قيمة مفقودة



-----------------------------------------------------------------------------

1. **Calculate basic statistics mean, median, mode, minimum, and maximum salary, determine the range of salaries, and find the standard deviation:**

لحساب الإحصائيات نقوم باستخدام التابع describe()



حيث أن :

* Count : عدد القيم الغير مفقودة في كل عمود
* mean : يقوم بحساب المتوسط الحسابي لكل عمود
* Std : الانحراف المعياري لكل عمود
* min : اصغر قيمة في كل عمود
* max : اكبر قيمة في كل عمود

-----------------------------------------------------------------------------

**3- Handle missing data by suitable method with explain why you use it:**

ﻧﻼﺣظ انه ﯾوﺟد اﻋﻣدة ﻻ ﺗﺣوي أي قيمة كالحقل notes وstatus و ﺣﻘل benefits يحوي ﻋﻠﻰ ﻗﯾم ﻓﺎرﻏﺔ ﺑﺎﻋداد ﻛﺑﯾرة جدا ﻟذﻟك نقوم ﺑﺣذﻓﮭم من خلال التابع drop

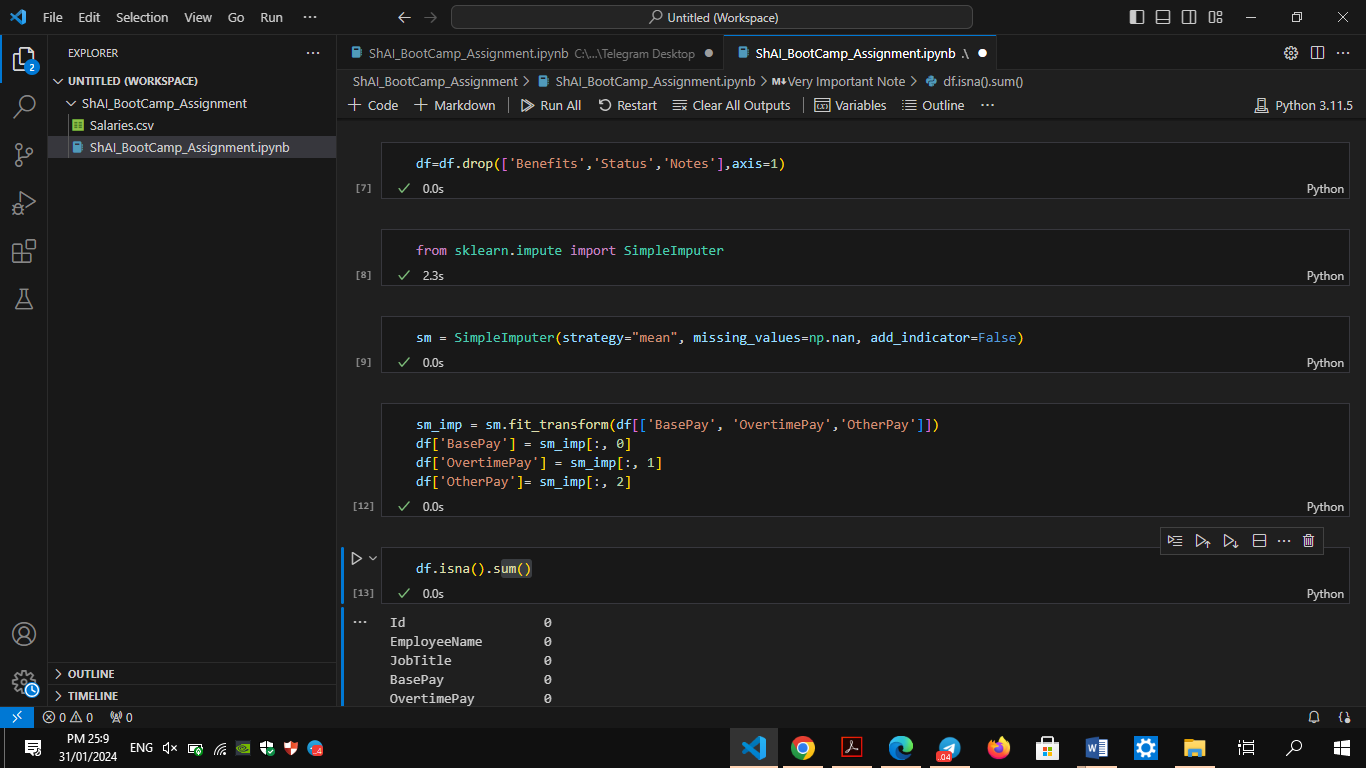
ثم نقوم بتعبئة القيم الفارغة التي قمنا بعرضها سابقاً من خلال احدى الطرق :

SimpleImputer

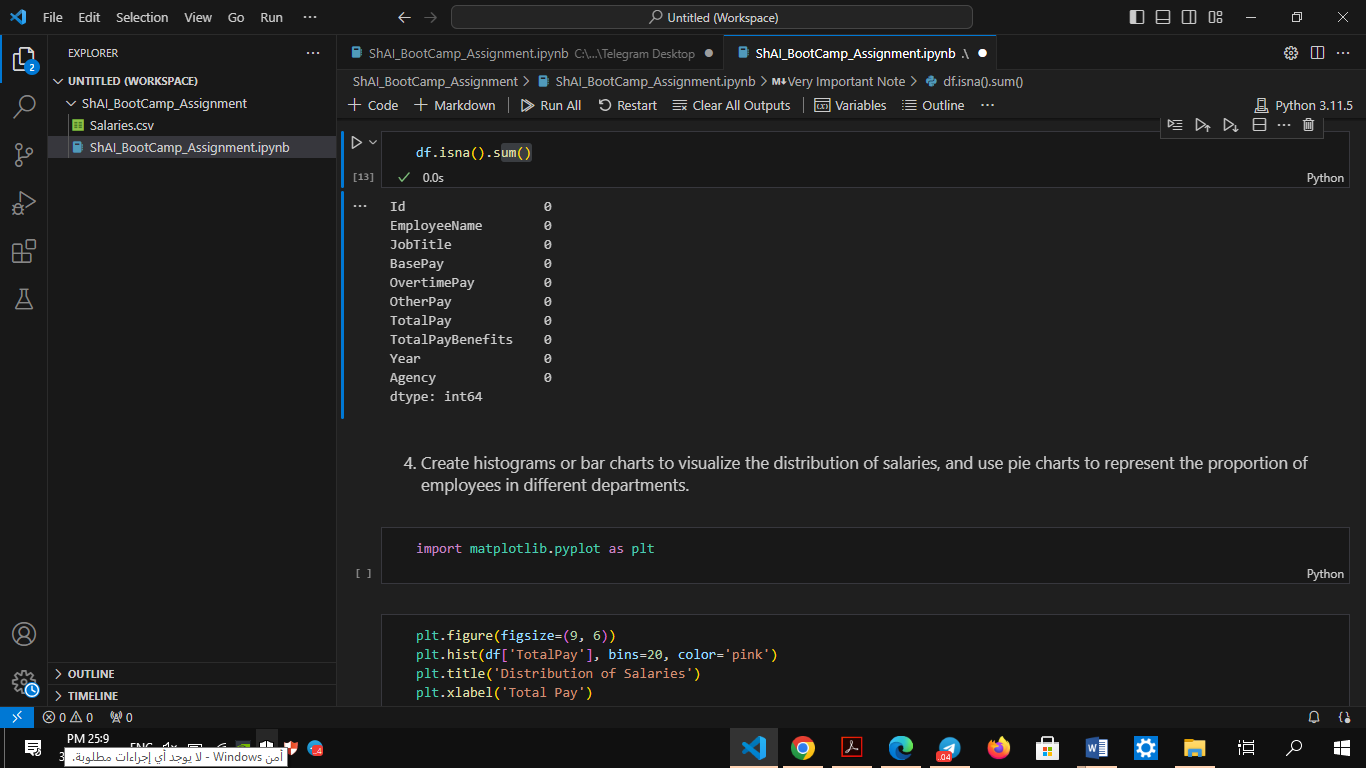
KnnImputer

تابع filna .

سنقوم باختيار SimpleImputer لتعبئة القيم الفارغة لان البيانات ذات طبيعة رقمية وذلك من خلال مكتبة sklearn.impute , ويتم تحديد القيم الافتراضية ل paramters مثلاً strategy = "mean", missing\_values = np.nan ) add\_indicator = False)



ثم نقوم بالتحقق من تعبئة كافة القيم المفقودة من خلال التابع isna ()



-----------------------------------------------------------------------------

**4- Create histograms or bar charts to visualize the distribution of salaries, and use pie charts to represent the proportion of employees in different departments:**

نقوم بإنشاء الرسوم البيانية من خلال مكتبة matplotlib.pyplot

