



Report TODO Task

DATASET SALARIES

❖ لقد تم تحليل التاسك وفهم الداتا سبت المدرجة وال features المتعلقة بها للبدء بحل طلبات التاسك الذي يعتمد على فهم أساسيات لغة البرمجة Python وتحليل البيانات . Data Science

- توصيف الداتا سبت:

DataSet Salaries: تحتوي على بيانات تعود ل148654 موظف بما يتعلق بطبيعة عملهم في شركة.

وتوفر 13 معلومة حول كل موظف مثل اسمه, طبيعة عمله, راتبه, القسم الذي يعمل به

والآن بما أن البيانات جاهزة لدينا قمنا بعملية معالجة للبيانات وفق ما يلي:

- ال cells الأساسية في ال notebook للحل:
(1)Cell: في البداية تم تحميل الداتاسبت وترتيب إظهار البيانات ضمن جدول يحوي على جميع الموظفين, والأعمدة التي تمثل خصائص كل موظف.
(2)Cell: للحصول على عدد الأعمدة.

Df: DataFrame هو الاسم الذي يحتوي على مجموعة البيانات التي نجري معالجة لها.

❖ حل التاسكات المطلوبة:

1- Basic Data Exploration:

تم تحديد عدد الصفوف والأعمدة في مجموعة البيانات، وتحديد أنواع البيانات في كل عمود، وفحص وجود القيم المفقودة في كل عمود.

تم استخدام تعليمة df.shape للحصول على عدد الأسطر والأعمدة.

وتعليمة df.dtypes لتحديد نوع البيانات لكل عمود.

وتم الكشف عن القيم الناقصة في البيانات باستخدام: df.isnull().sum().

تم استخدام تعليمة df.axes للحصول على محاور البيانات في مجموعة البيانات, حيث باستخدام هذه المعلومات يمكن الوصول إلى الصفوف والأعمدة في مجموعة البيانات والقيام بالعديد من العمليات الأخرى مثل تحديد الصفوف أو الأعمدة المحددة.

2- Descriptive Statistics:

تم حساب بعض الإحصائيات الأساسية للرواتب

3- Data Cleaning:

تم التعامل مع وجود بيانات ناقصة, في تعليمة سابقة اختبرنا وجود بيانات ناقصة حيث يمكن التعامل مع هذه المشكلة بأكثر من طريقة:

-حذف البيانات الناقصة: تم استخدام هذه الطريقة باعتبار أن البيانات الناقصة قليلة بالنسبة لحجم البيانات الكلي.

-إدخال البيانات الناقصة: يمكن استخدام هذه الطريقة باستخدام قيم معينة مثل الوسيط أو المتوسط.

4- Basic Data Visualization:

5- Grouped Analysis:

6- Simple Correlation Analysis: