



SHAI ASSIGIMENT

Shaimaa AlRajeh



JANUARY 27, 2024

SHAI COMPANY

الخطوة الأولى: استكشاف البيانات.

في الخطوة الأولى من العمل قمنا بعمل استكشاف للبيانات من خلال عدة خطوات وذلك من أجل أخذ نظرة عامة عن البيانات وكيفية التعامل معها :

1. قمنا باستيراد المكاتب والبيانات وعرضها .
2. معرفة أسماء الأعمدة , وحجم البيانات, معرفة الأعمدة التي تحوي قيم فارغة من أجل معالجتها سابقاً .

الخطوة الثانية : الوصف الإحصائي .

في هذه الخطوة قمنا بعمل بعض الإحصائيات على البيانات الراتب ومن ثم قمنا بتحديد حدود الرواتب , وهنا قمنا بأخذ القيم الموجبة تماماً , والتي تمثل الرواتب, حيث لا يمكن أن يكون إجمالي الراتب سالب , حيث أن الفوائد لا يمكن أن تغطي أكثر من إجمالي الراتب .

الخطوة الثالثة: تنضيف البيانات .

قمنا بملى ال قيم الفارغة بالرقم 0 , حيث أن الأعمدة التي تحتوي قيم فارغة لا يمكن أن نضع فيها المتوسط او القيم اسابقة او اللاحقة كون كل قيمة مستقلة عن الأخرى, فعدم وجود أحد القيم يعني أنها معدومة , وبالنسبة للأعمدة (Notes Status) فهي أعمدة جميعها قيم فارغة ولا تفيد بأي معلومة مهمة أثناء العمل , قمنا بتصغيرها أيضاً , ويمكننا حذف العمودين بشكل كامل دون التأثير على أي نتائج بالعمل .

قمنا بعدها بالتأكد من عدم وجود قيم فارغة.

الخطوة الرابعة: رسومات أساسية .

مخطط ال Histogram من أجل أخذ نظرة عن إجمالي الرواتب ومعدل تكرارها , قمنا باستنتاج متوسط الرواتب , وهو أعلى ارتفاع ضمن المخطط .

من أجل رسم المخطط الآخر والذي من خلاله يجب معرفة نسب الموظفين بكل قسم, أولاً قمنا بمعرفة عدد الموظفين من خلال أخذ عدد الموظفين بكل قسم من الأقسام ثم قمنا بتمثيل هذا العدد, هذا الرسم البياني قام بإعطائنا نظرة شاملة وخاطفة حول الأقسام ذات عدد الموظفين الأكثر .

الخطوة الخامسة: تحليل المجموعات .

هنا قمنا بتجميع البيانات حسب السنة ونوع الوظيفة, والسبب بذلك أنني أريد معرفة أكثر الوظائف رواتب سنوياً.

قمنا بأخذ متوسط الرواتب بعد تجميعها , قمنا بعد لك برسمها من أجل أخذ نظرة أكثر قرباً للنتائج.
وقمنا بمعرفة أكثر وأقل الوظائف رواتب .

الخطوة السادسة: تحليل الترابط .

من التحليل البسيط الذي قمنا فيه استنتجنا أن العلاقة بين إجمالي الراتب والراتب الأساسي علاقة ترابط قوية, أي أن الفوائد والاضافات ... لا تؤثر كثيراً على الراتب سواء بالزيادة أو النقصان.

ومن المخطط المبعثر نرى أن العلاقة شبه خطية أي أنها قوية بشكل جيد.