

بعض الاستنتاجات من dataset التي تم العمل عليها:

- عدد الاعمدة الكلي هو 13 وعدد الاسطر الكلي هو 148654.
- الاعمدة التي تحوي على قيم فارغة هي BasePay, OvertimePay, OtherPay, Benefits, Notes and Status.
- في عملية تنظيف الداتا تم استخدام median الخاص بكل عامود لوضع هذه القيمة في القيم الفارغة او المفقودة ولم استخدم mean وذلك لوجود فارق كبير بين كلا من ال mean و median وايضا قمت بعملية حذف لكلا من العامودين Status and Notes وذلك لانهم يحتوا قيمة Nan وبالتالي لا يمكن الاستفاد منهم وحتى العامود الخاص ب Agency يحوي فقط على قيمة واحد وهي San Francesco وبالتالي يمكن حذفها ايضا.
- في عملية عرض ال Pie Chart الخاص بالاقسام لقد قمت بالعديد من عمليات التوحيد حتى اتمكن من رسم Pie Chart كثر دقة ولكن يوجد الكثير من العنوانين التي لم اتمكن من ان احدد لاي قسم تنتمي وفي الغالب اسخراج الفيتشر هذه بدقة اكبر ووضح تتطلب الكثير من العمل و التنسيق الاكبر مع مصدر الداتا حتى تكون الفيتشر هذه واقعية وحيوية اكثر ولقد قمت بعد عمليات التوحيد البسيطة التي قمت بها باخذ فقط الاقسام العشرة التي تحوي اكبر عدد موظفين بناءا على عملية التوحيد التي قمت بها.
- في عمليات تجميع الداتا بناءا على قيمة فيتشر ما لقد اعتمدت على Year ومن ثم وجدت ان متوسط الرواتب في عام 2013 هو الكبر وقيمه 77611.44 بينما في 2011 هو الاصغر قيمته 71744.10
- في تحديد الارتباط بين ال TotalPay وباقي الفيتشر وجدت علاقة ارتباط بينها وبين كلا من BasePay and TotalPayBenefits