**Objective:**

The goal is to predict whether the salary for data-related positions, based on location, company reviews, and job title, is satisfactory or not.

**هدف:**

در این پروژه یک مجموعه داده از حقوق مشاغل مختلف که با داده سروکار دارن (مثل دیتا ساینتیست، تحلیلگر داده و ...) در ایالت‌های مختلف آمریکا داریم و مبلغ حقوق به اینکه کدوم ایالته، بقیه در مورد شرکت چی گفتن، و اینکه عنوان دقیق شغل چیه و... بستگی داره. حالا هدف ما اینه که بفهمیم (پیش‌بینی کنیم) حقوق این مشاغل مناسبه یا نه.

وقتی به این داده نگاه می‌کنیم، چند تا چیز به چشم می‌خوره:

ما می‌تونیم عناوین شغلیو براساس کلماتی که در اونها وجود داره دسته بندی کنیم: مثلاً همه کارهایی که تو اسمشون کلمه

"Data Scientist" یا "Data Science" هست رو بذاریم یه جا.

کمترین و بیشترین حقوق رو از اون ستون "Salary Estimate" استخراج کنیم، میانگین بگیریم و بگیم که این، حقوق متوسط کارهاست. و در نهایت این متغیر هدف ما خواهد بود.

فکر نکنم اسم شرکت‌ها خیلی فرقی داشته باشه.

از ستون "Location" می‌تونیم بفهمیم هر کار کجا انجام میشه.

ستون "Revenue" خیلی اطلاعاتش دقیق نیست، پس فعلاً کاری بهش نداریم