

# Job Salary Prediction

Inzamamul Alam

## THE DATASET

---

My task for this assignment is **Job Salary Prediction**. This is a competition on Kaggle. The data I choose is a set of job ads published in the UK. The Description of the data sets are

**Id** - A unique identifier for each job ad

**Title** - A freetext field supplied to us by the job advertiser as the Title of the job ad. Normally this is a summary of the job title or role.

**FullDescription** - The full text of the job ad as provided by the job advertiser. Where you see \*\*\*, we have stripped values from the description in order to ensure that no salary information appears within the descriptions. There may be some collateral damage here where we have also removed other numerics.

**LocationRaw** - The freetext location as provided by the job advertiser.

**LocationNormalized** - Adzuna's normalised location from within our own location tree, interpreted by us based on the raw location. Our normaliser is not perfect!

**ContractType** - full\_time or part\_time, interpreted by Adzuna from description or a specific additional field we received from the advertiser.

**ContractTime** - permanent or contract, interpreted by Adzuna from description or a specific additional field we received from the advertiser.

**Company** - the name of the employer as supplied to us by the job advertiser.

**Category** - which of 30 standard job categories this ad fits into, inferred in a very messy way based on the source the ad came from. We know there is a lot of noise and error in this field.

**SalaryRaw** - the freetext salary field we received in the job advert from the advertiser.

**SalaryNormalised** - the annualised salary interpreted by Adzuna from the raw salary. Note that this is always a single value based on the midpoint of any range found in the raw salary. This is the value we are trying to predict.

**SourceName** - the name of the website or advertiser from whom we received the job advert.

The dataset contains 244768 rows. 75% of the data was chosen at random and was used for training while the rest was kept for testing. There is nowhere mentioned if the salaries are in US dollar or pounds so it is referred as units per annum. There is data for about 2732 places and there are 135,436 unique Jobs titles present. The main categories that the jobs fall under are given below with their average salaries next to them

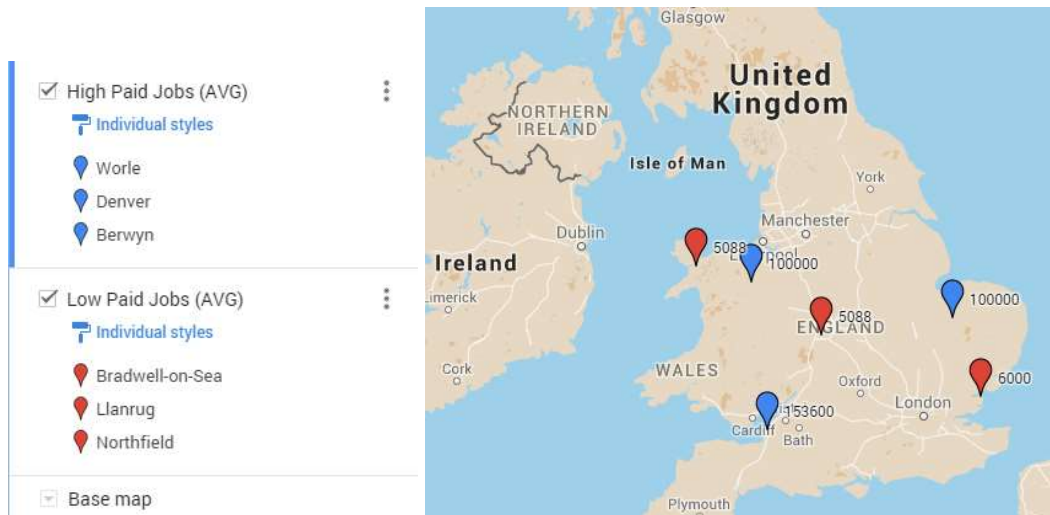
Accounting & Finance Jobs	38622.456172
Admin Jobs	20916.362130
Charity & Voluntary Jobs	28200.204936
Consultancy Jobs	36374.208544
Creative & Design Jobs	32585.487409
Customer Services Jobs	19795.438648
Domestic help & Cleaning Jobs	17492.815789

<b>Energy, Oil &amp; Gas Jobs</b>	<b>45384.110103 (Max)</b>
Engineering Jobs	35608.058499
Graduate Jobs	28677.438703
HR & Recruitment Jobs	32386.298070
Healthcare & Nursing Jobs	32203.188169
Hospitality & Catering Jobs	23552.462798
IT Jobs	44081.780990
Legal Jobs	42350.550210
Logistics & Warehouse Jobs	25711.287983
Maintenance Jobs	17533.320433
Manufacturing Jobs	25653.276129
Other/General Jobs	34361.716029
PR, Advertising & Marketing Jobs	35521.313285
<b>Part time Jobs</b>	<b>10514.285714 (Min)</b>
Property Jobs	32167.056647
Retail Jobs	32851.304998
Sales Jobs	30685.809207
Scientific & QA Jobs	33950.363683
Social work Jobs	32324.809030
Teaching Jobs	27240.637182
Trade & Construction Jobs	36050.166566
Travel Jobs	23913.173143

The average salary in UK is 31,554.441 units. The maximum salary is in 'Worle' and 'Denver' with the value being 153,600 and 100,000 respectively. However, the minimum salary being in 'Northfield' and 'Llanrug' which is just 5,088 units. One surprising finding was that the average salary in 'Channel Islands' is about 71,500 which is about 40,000 above average; however, the rest of the places are about 10,000 above or below average. Channel Islands is not one of LocationNormalised feature in the data. There is another file given which gives a tree like structure of the places in UK. Something like Country~State~city. The comparison of Channel Islands is with other states. Although, Worle, Denver etc are at city level.

Another major finding was that the maximum standard deviation among the salaries based on Job titles was found in 'Reconciliations Accountant' with the deviation being 85,920. The maximum salary for this job title is 199,860 and minimum is 27,840. Although, the person who earns the most in UK has a salary of 200,000 and works in an Accounting & Finance Job.

This dataset also tells if people work on contract or are permanent. The initial instinct is that the people who work on contract would be paid more because they do not have any job security and in the USA the freelancers get paid higher than permanent workers. Surprisingly, there is not much difference in the income of the people on contracts and permanent employee in UK. People on contract earn about 36,099.67 where as permanent employee get around 35,327.47 which is just a difference of 700.



Map1: Cities with highest and lowest average salaries

Map 1 shows the plot of the cities with highest and lowest average salary. The blue plots are highest salary and the red are lowest average salary at city level. Besides each plot is the average salary mentioned. There is no clear trend based on the geographical locations for the salaries earned. However, London and the South East have a higher income at average.