Salary Prediction EDA

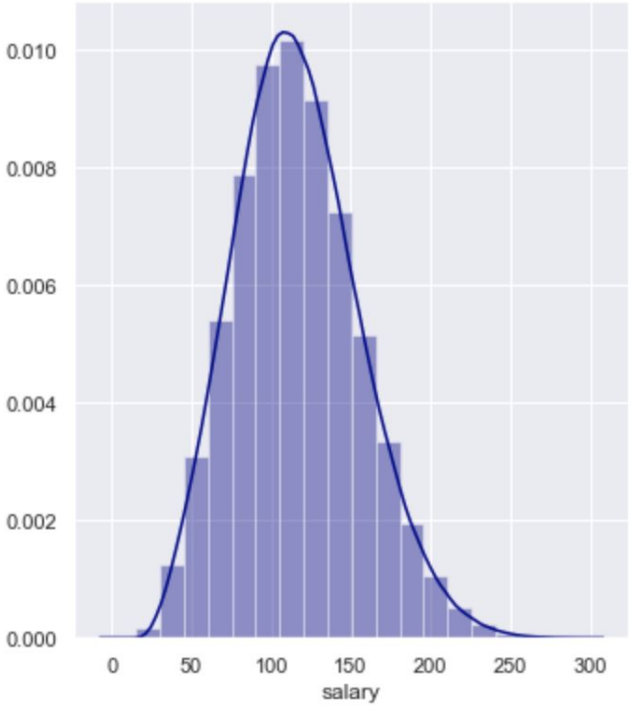
Alena Nazarava

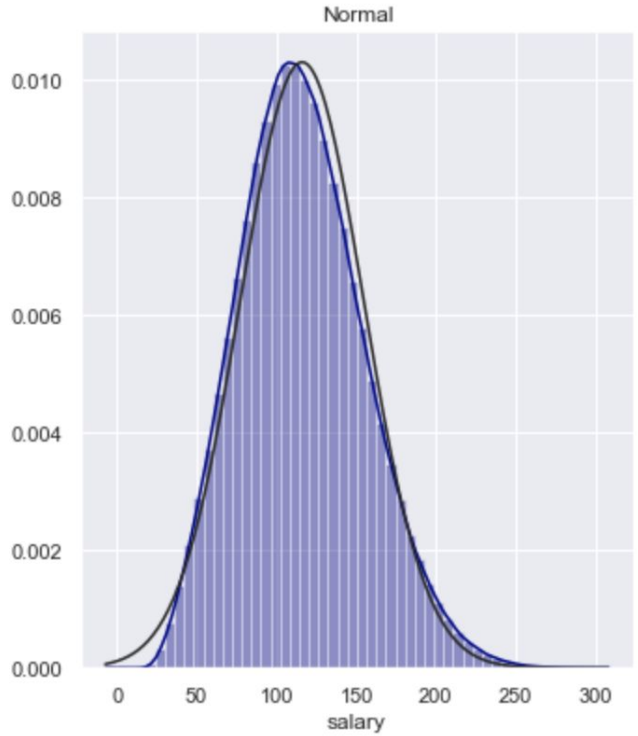
Contents

The goal is to predict the salary of a job postings based on the given information. Since we are predicting a continuous value of salary for particular Job, the formulation of the salary prediction problem will be defined as a regression task.

- Convert data into useful information.
- Examine the distribution of the variable and relationship between features.
- Explore which features have an influence on salaries, and how they may be related.
- Using this analysis, the approach to the problem can be determined.

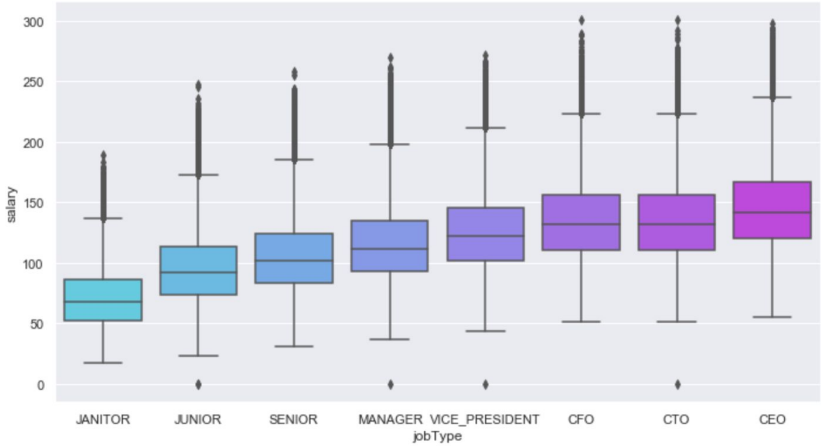
Salary Distribution





Salary tends to follow a **right-skewed** distribution

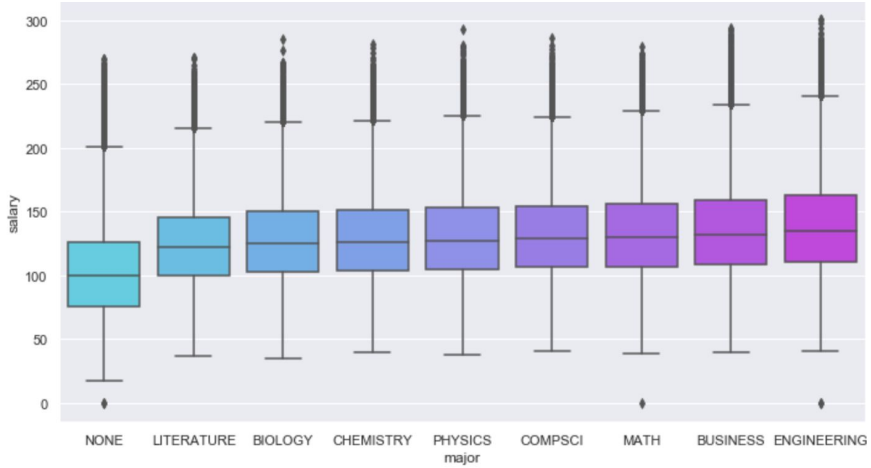
JobType and Salary



jobType	count	mean	std	min	25%	50%	75%	max
CEO	124778.0	145.311425	34.423846	55.0	120.0	142.0	167.00	298.0
CFO	124369.0	135.458547	33.069203	51.0	111.0	132.0	156.00	301.0
CTO	125046.0	135.479983	33.093032	0.0	111.0	132.0	156.00	301.0
JANITOR	124971.0	70.813045	24.643021	17.0	52.0	68.0	86.00	189.0
JUNIOR	124594.0	95.331557	29.249018	0.0	74.0	92.0	113.75	248.0
MANAGER	125121.0	115.367596	30.835467	0.0	93.0	112.0	135.00	270.0
SENIOR	125886.0	105.487775	30.069722	31.0	83.0	102.0	124.00	258.0
VICE_PRESIDENT	125235.0	125.367629	31.958131	0.0	102.0	122.0	146.00	272.0

JANITOR jobs, as a group, have less salary than others. The median number of salary for **JANITOR job (68)** is less than the median (and even the first quartile) of either of the other distributions.

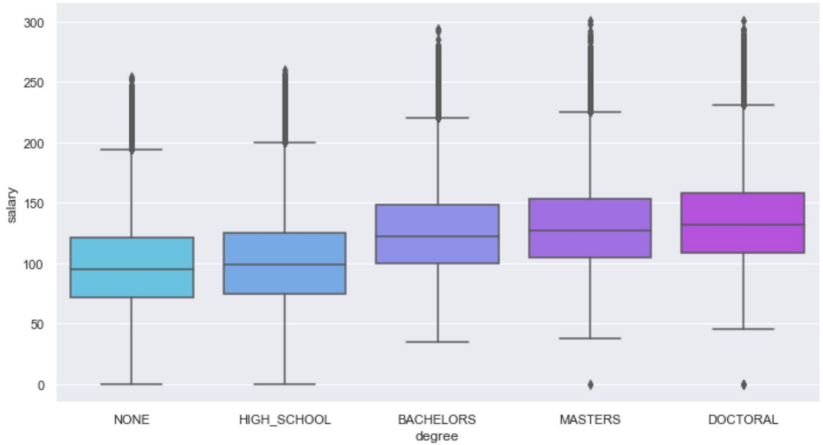
Major and Salary



major	count	mean	std	min	25%	50%	75%	max
BIOLOGY	58379.0	127.932664	34.820141	35.0	103.0	125.0	150.0	285.0
BUSINESS	58518.0	135.648980	37.006526	40.0	109.0	132.0	159.0	294.0
CHEMISTRY	58875.0	129.072085	34.832396	40.0	104.0	126.0	151.0	282.0
COMPSCI	58382.0	132.075605	34.776640	41.0	107.0	129.0	154.0	286.0
ENGINEERING	58596.0	138.436617	38.186849	0.0	111.0	135.0	163.0	301.0
LITERATURE	58684.0	124.423097	33.576413	37.0	100.0	122.0	146.0	271.0
MATH	57801.0	133.319735	35.543265	0.0	107.0	130.0	156.0	280.0
NONE	532355.0	102.583479	36.136748	0.0	76.0	100.0	126.0	270.0
PHYSICS	58410.0	130.372436	34.956981	38.0	105.0	127.0	153.0	293.0

NONE major has lower salary than other majors. The median number of **NONE major (100)** is less than the median of other distributions (median LITERATURE: 122, MATH: 130). **NONE IQR is 50** and **LITERATURE IQR is 46**, but the full ranges don't vary a lot. In general NONE major showed worse results and has a large variety of values.

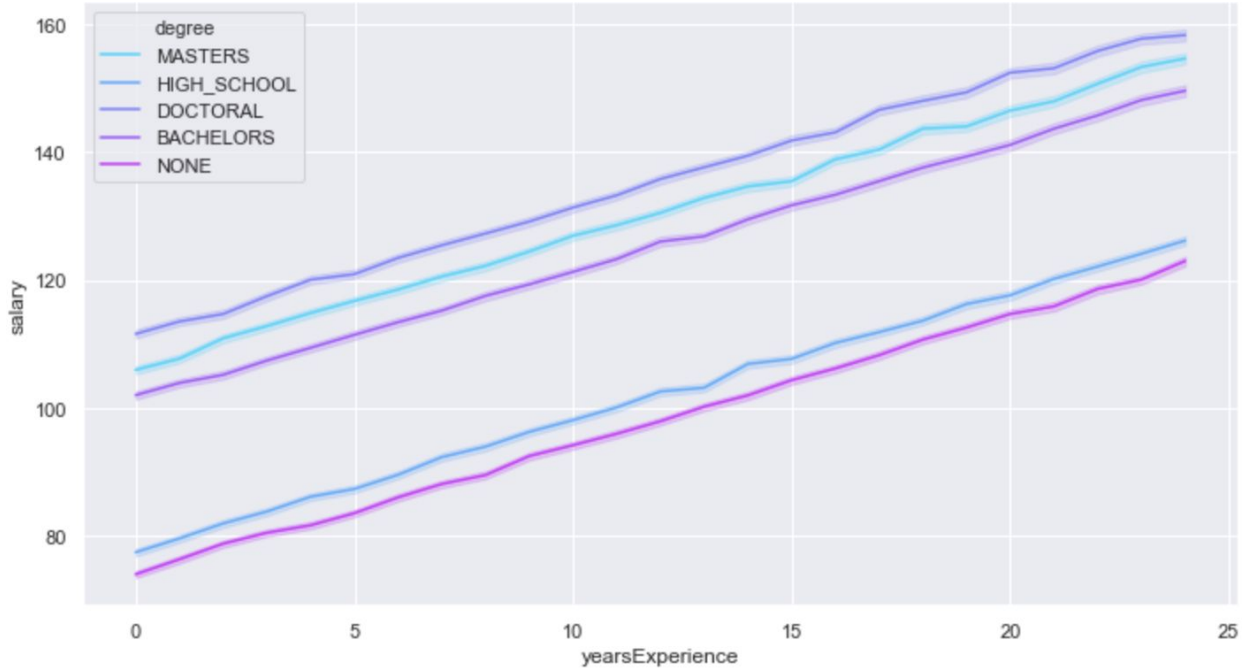
Degree and Salary



degree	count	mean	std	min	25%	50%	75%	max
BACHELORS	175495.0	125.454663	35.042720	35.0	100.0	122.0	148.0	294.0
DOCTORAL	175364.0	135.489433	35.793276	0.0	109.0	132.0	158.0	301.0
HIGH_SCHOOL	236976.0	101.920654	36.113215	0.0	75.0	99.0	125.0	260.0
MASTERS	175311.0	130.504903	35.401101	0.0	105.0	127.0	153.0	301.0
NONE	236854.0	98.176467	34.915650	0.0	72.0	95.0	121.0	254.0

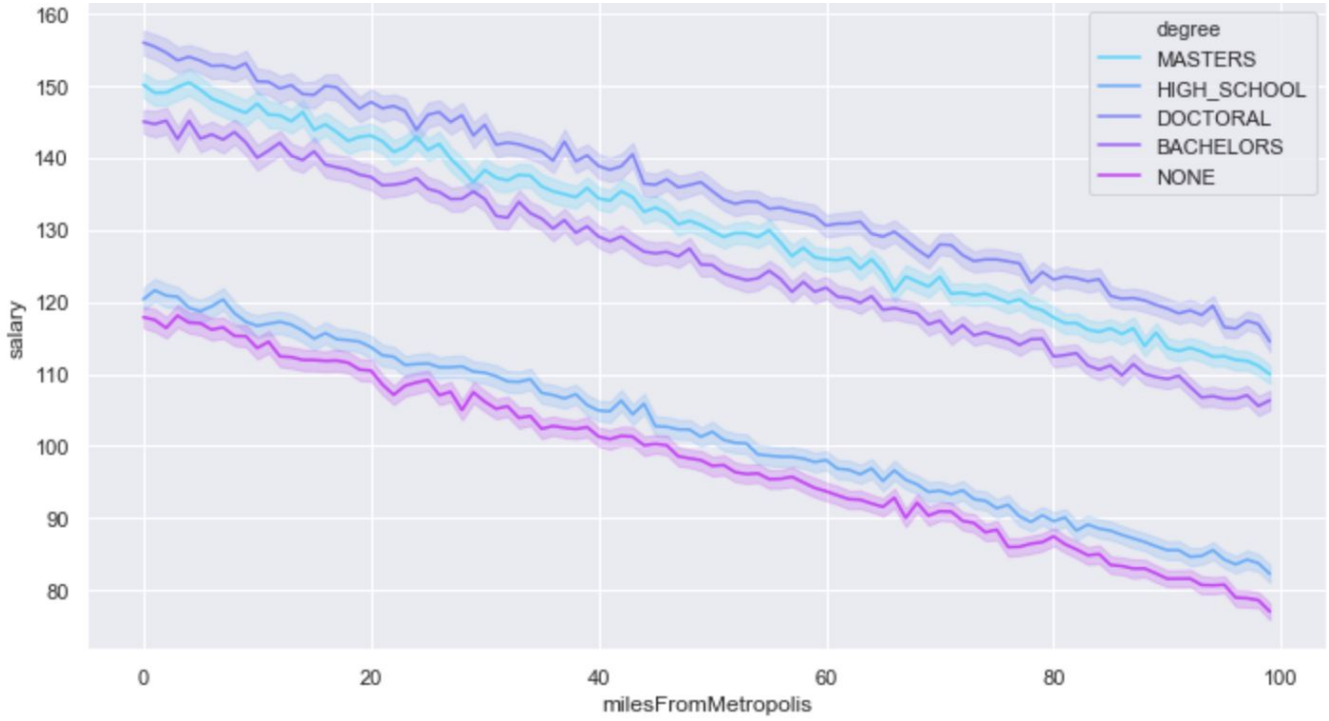
NONE and **HIGH_SCHOOL** have worse salaries. The median number of **NONE** and **HIGH_SCHOOL** degree **(95, 99)** is less than the median of other distributions (median BACHELORS: 122, MASTERS: 127, DOCTORAL: 132)

YearsExperience and Salary



Individuals with **more** years of experience have **higher** salaries.

MilesFromMetropolis and Salary



Individuals who work **further** from Metropolis have **lower** salaries

How **degree** is related to **jobType?**

degree jobType	BACHELORS	DOCTORAL	HIGH_SCHOOL	MASTERS	NONE	Total
CEO	50003	50298	49708	49755	49955	249719
CFO	49887	49940	49950	50088	49596	249461
CTO	49904	49969	49778	50401	49659	249711
JANITOR	0	0	125129	0	125095	250224
JUNIOR	49583	50074	50280	49922	49757	249616
MANAGER	49758	49776	50055	49970	49953	249512
SENIOR	50394	50182	49902	50249	50361	251088
VICE_PRESIDENT	49903	50230	50429	50162	49945	250669
Total	349432	350469	475231	350547	474321	2000000

JANITOR jobType has **50%** HIGH_SCHOOL degree and **50%** NONE degree.