SALARIES IN TECH:

AI VS DATA SCIENCE

Weston Slayton, Sophia Monaco,
Elliot Beamer, & Amelia Lane

The Topic and why we chose it

- Research question: Salaries in AI vs data analyst/scientist Jobs, change over time
- Difference between US and non US
- Hypothesis: AI salaries are higher and increasing rapidly
- US salaries higher

- Intrigued by the introduction of AI into our everyday lives (ChatGPT)
- Wanted to see if AI was more lucrative compared to other tech jobs
- Interesting analysis of the post-COVID economic landscape in tech
- Data science class, why not think about data science!

Introducing the Data

- Data is from ai-jobs.net, posts tech jobs and analyzes trends
- Each observation records individual job info (see variables on the right!)
- Ethical issue: self-reported data
- For our analysis: combined ML Engineer with AI Jobs into variable ML/AI, and combined data scientists with data analysts into variable datascientistanalyst

Work year

Job title

Salary

Employee residence

Company location

Remote ratio

Experience level

Company size

Employment type

Hypothesis Test/CLT

```r
#| label: observed stat

salary_jobs |>
  group_by(job_type) |>
  summarise(mean_salary = mean(salary_in_usd))
```

A tibble: 2 × 2

job_type <chr>	mean_salary <dbl>
Data Scientist/Ana…	123528.2
ML/AI	147410.2

2 rows

```r
#| label: CLT-check

salary_jobs |>
  group_by(job_type) |>
  summarise(n = n())
```

A tibble: 2 × 2

job_type <chr>	n <int>
Data Scientist/Ana…	1173
ML/AI	296

```r
#| label: CLT

salary_jobs |>
  group_by(job_type) |>
  summarise(sd_salary = sd(salary_in_usd))

sdAI <- 64176.44
sdData <- 57389.76
nAI <- 296
nData <- 1173
SE <- sqrt(((sdAI*sdAI)/(nAI)) + ((sdData*sdData)/(nData)))
```

A tibble: 2 × 2

job_type <fctr>	sd_salary <dbl>
Data Scientist/Ana…	57389.76
ML/AI	64176.44

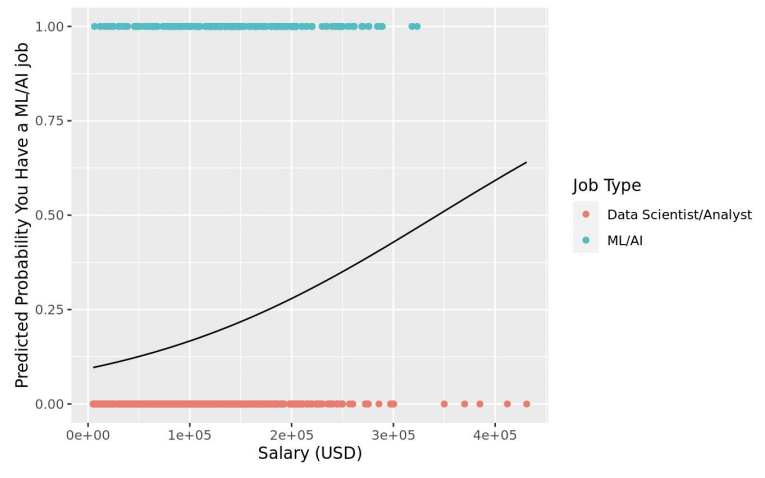
$$\bar{x}_{ai} - \bar{x}_{data} = 23882$$

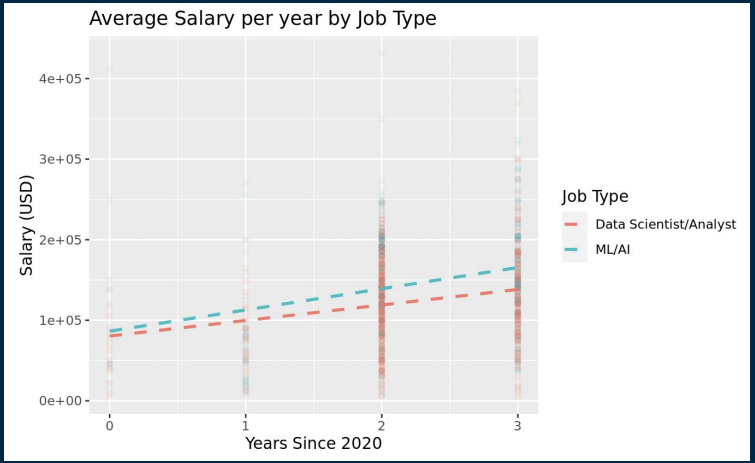
```r
#| label: p-value

tscore <- (23882 - 0) / SE

pt(tscore, df = 295, lower.tail = FALSE)
```

```
[1] 6.877573e-09
```

- Calculated observed statistic, mean salary
- Used CLT and calculated SE
- Found p value
- Rejected null hypothesis

Regression Models





A tibble: 4 × 5

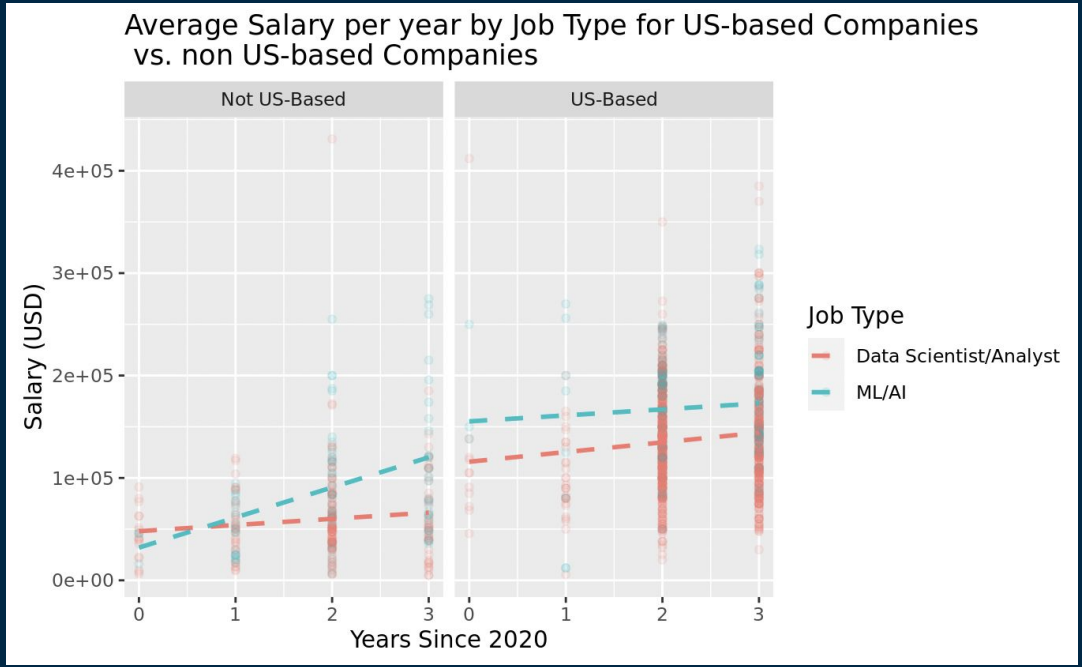
term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>
(Intercept)	80463....	5865.556	13.7179...	2.212746e-...
work_year	19258....	2514.860	7.6579193	3.412890e-...
job_typeML/AI	5986.455	12608....	0.4747845	6.350113e-...
work_year:job_typeML/AI	7083.503	5246.814	1.3500578	1.772060e-...

$$\widehat{salary\_in\_usd} = 80463.128 + 19258.594 * year + 5986.455 * ML/AI + 7083.503 * year * ML/AI$$

Data Scientist/Analyst: $\widehat{salary\_in\_usd} = 80463.128 + 19258.594 * year$

ML/AI: $\widehat{salary\_in\_usd} = (80463.128 + 5986.455) + (19258.594 + 7083.503) * year$

Faceted Visualization



Conclusion

- AI/ML salaries are comparatively higher
- Overall AI/ML salaries growing at a faster rate
- In the US, AI/ML salaries are growing slower than in non US countries and slower than Data scientist salaries
- Important to note some possible inaccuracies with data

Thank You!