

DATA PROCESSING

01 CLEANING PROCESS

- Clean job description
- Transform “*salary*” and “*size*” and “*sector*” columns to dummy variables
- NLP pipeline

03 TOPIC MODELLING

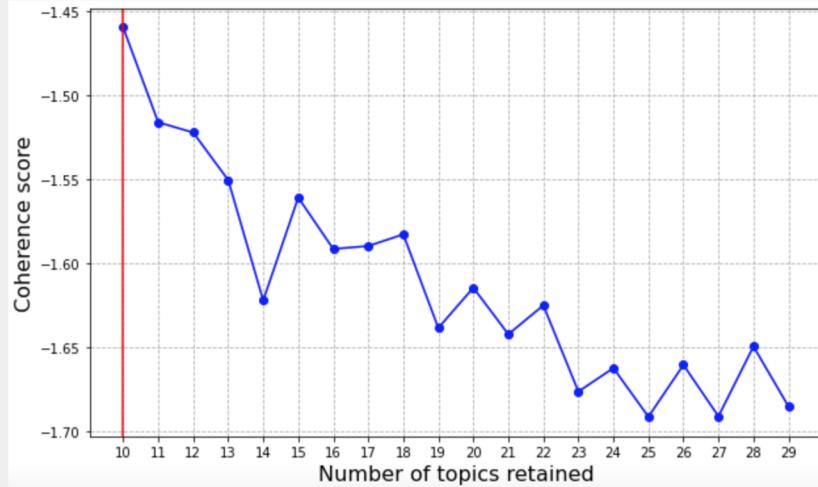
- Calculate coherence values
- Build the model with 10 topics
- Explore insights of the topics
- Multidimensional Scaling Analysis

02 APPLICATION OF TF-IDF

Filter out less meaningful words in each documents by removing tokens that have less than 0.02 TF-IDF value

04 INTERESTING FINDING

The requirements change across job title, salary range, and sectors



Coherence values correspond to the number of topics retained

TOPIC MODELLING

The terms reflect the relationship between job and sector/ industry/ experience/ and skill requirement

	0	1	2	3	4	5	6	7	8	9
Topic_1	position	information	application	program	support	system	require	service	provide	education
	(0.026)	(0.019)	(0.018)	(0.016)	(0.016)	(0.016)	(0.013)	(0.01)	(0.009)	(0.009)
Topic_2	financial	security	status	disability	risk	service	employment	contract	employee	provide
	(0.023)	(0.017)	(0.012)	(0.011)	(0.01)	(0.009)	(0.009)	(0.008)	(0.008)	(0.007)
Topic_3	ability	report	ensure	process	management	Maintain	quality	reporting	information	require
	(0.036)	(0.029)	(0.018)	(0.017)	(0.016)	(0.015)	(0.014)	(0.011)	(0.011)	(0.01)
Topic_4	knowledge	skill	strong	database	sql	system	report	analyst	management	query
	(0.042)	(0.039)	(0.038)	(0.032)	(0.029)	(0.023)	(0.021)	(0.019)	(0.018)	(0.018)
Topic_5	business	analysis	analytic	tool	develop	visualization	dashboard	tableau	analytical	model
	(0.051)	(0.042)	(0.041)	(0.026)	(0.016)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)
Topic_6	product	build	marketing	team	insight	company	people	role	analytic	engineering
	(0.024)	(0.015)	(0.015)	(0.014)	(0.013)	(0.012)	(0.01)	(0.008)	(0.007)	(0.006)
Topic_7	team	support	product	partner	operation	organization	sale	strategy	lead	service
	(0.051)	(0.016)	(0.016)	(0.016)	(0.015)	(0.014)	(0.013)	(0.012)	(0.012)	(0.011)
Topic_8	client	company	opportunity	technology	customer	good	professional	solution	industry	environment
	(0.044)	(0.019)	(0.016)	(0.014)	(0.013)	(0.012)	(0.01)	(0.01)	(0.009)	(0.008)
Topic_9	research	health	analysis	healthcare	statistical	clinical	science	care	program	medical
	(0.03)	(0.021)	(0.017)	(0.017)	(0.012)	(0.01)	(0.01)	(0.01)	(0.009)	(0.008)
Topic_10	business	development	project	process	technical	management	solution	system	quality	design
	(0.048)	(0.025)	(0.024)	(0.023)	(0.019)	(0.018)	(0.018)	(0.017)	(0.017)	(0.013)

Financial Data Analyst (Topic 2)

Data analyst in Data Management System (Topic 4)

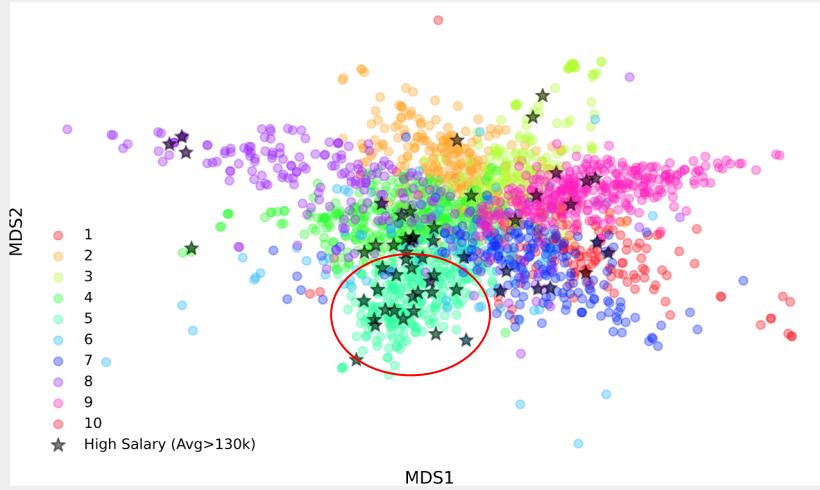
Data analyst with strong data visualization skills (Topic 5)

Senior Business Analyst (Topic 7)

Data analyst in Health Care Service & Hospitals industry (Topic 9)

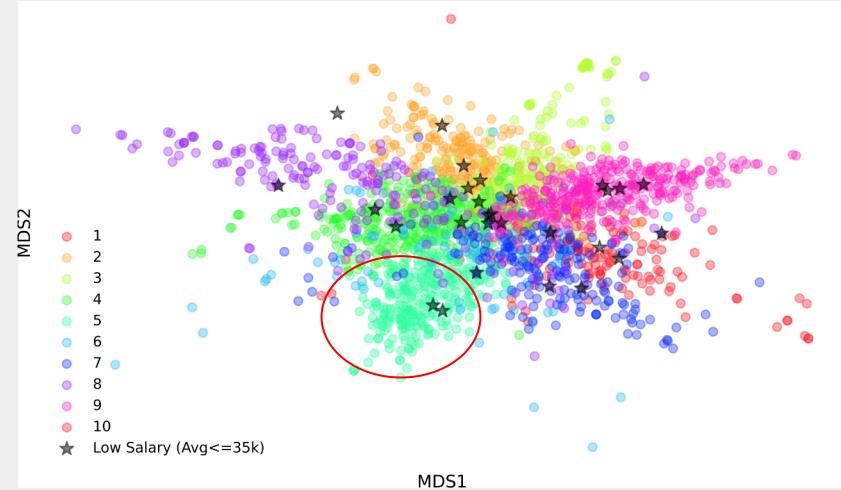


TOPIC MODELLING



High salary jobs are highly related to **Topic 4 and 5** which require strong skills of data visualisation and data management

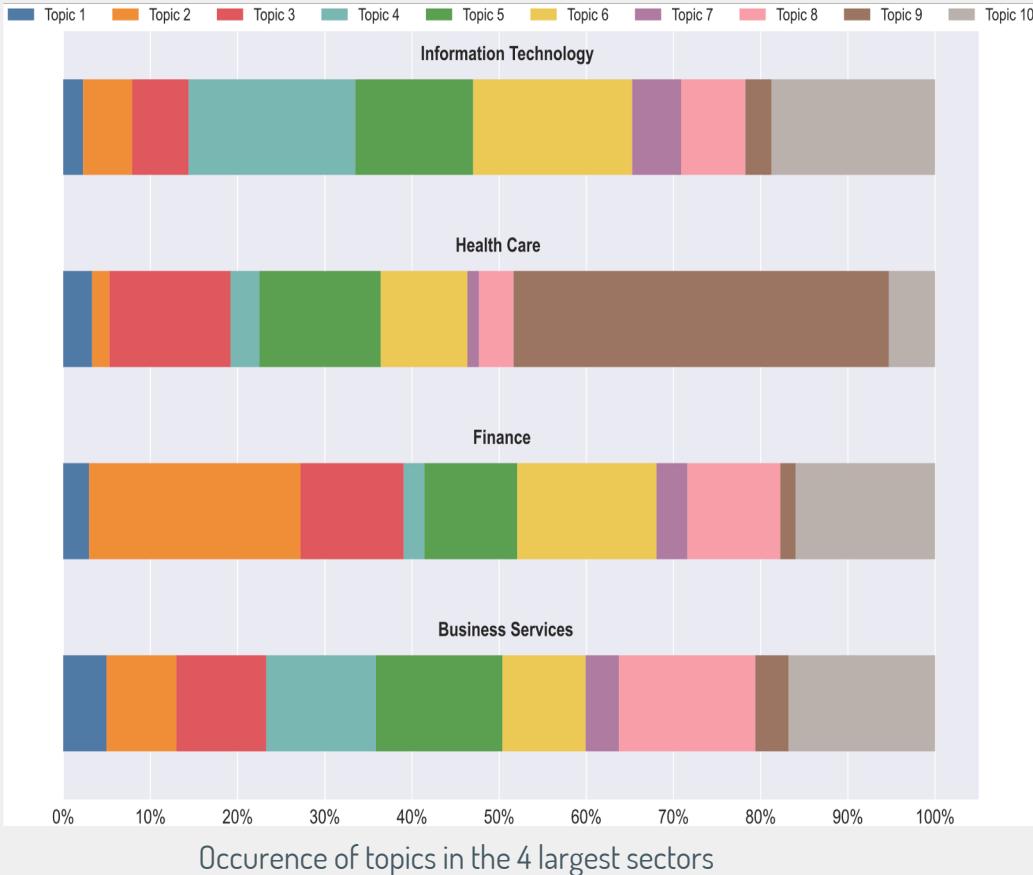
 **Important terms:** "sql", "database", "query", "dashboard", "visualization", "tableau", "strong"



- * Low salary jobs exist in all topics generally, which may indicate the similarity level of salary for data analysts at entry level across sectors and industries;
- * Compared to high salary, low salary jobs scarcely appear in Topic 5

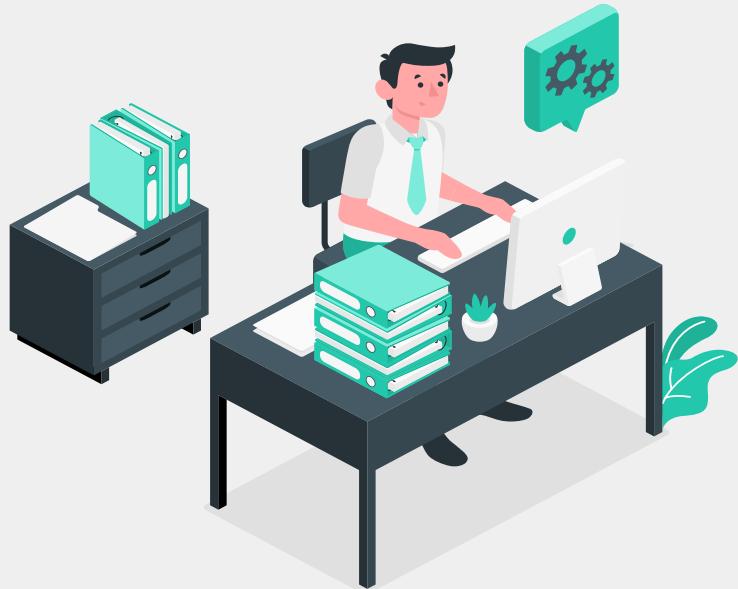
 **Top 2 terms for Topic 5:** "business", "analysis"

INTERESTING INSIGHTS



- Health Care sector is represented by a large proportion of **Topic 9** ("research", "health")
- Finance sector is highly associated with **Topic 2** ("financial", "security", "risk")
- **Topic 1 (potentially graduate)** and **Topic 7 (senior role)** have little contribution in all the 4 sectors
- **Topic 5 ("business", "analysis", "tool")** has similar proportion of contribution in all the 4 sectors

CONCLUSION



POINT 1

People having a combination of business knowledge and analysis skills are likely to earn high salary

POINT 2

People with skill sets in Topic 4 are highly recommended to find jobs in Information Technology sector

POINT 3

Data visualization skills are important as they are proportionally demanded across all the big sectors