



Analisa Pekerjaan Sains Data di Australia menggunakan Pendekatan Exploratory Data Analysis

Kartika Maulida Hindrayani¹

¹Sains Data UPN “Veteran” Jawa Timur

¹kartika.maulida.ds@upnjatim.ac.id

Abstract: Exploratory Data Analysis (EDA) is a field of data analysis used to visually represent the knowledge embedded deep in the given data set. Dataset of a data scientist job in Australia is currently shared in public. Data visualization is implemented to the dataset to give better insight about the topic.

Keywords: type your keywords here...3-5 keywords

Abstrak: Analisis Data Eksplorasi (EDA) adalah bidang analisis data yang digunakan untuk mewakili pengetahuan yang tertanam jauh di dalam kumpulan data yang diberikan secara visual. Kumpulan data pekerjaan ilmuwan data di Australia saat ini dibagikan ke publik. Visualisasi data diimplementasikan ke dataset untuk memberikan wawasan yang lebih baik tentang topik tersebut.

Kata kunci: Analisis Data, Pekerjaan Sains Data, EDA

I. PENDAHULUAN

Saat ini sains data merupakan ilmu yang banyak didengungkan dan populer tidak hanya secara nasional tetapi juga secara internasional. Sains data merupakan ilmu yang dikembangkan dengan perpaduan dari statistika/matematika, komputasi, dan domain bisnis. Kepopuleran sains data seiring dengan kemajuan dari *Big Data*, sehingga dibutuhkan lulusan atau profesi yang dapat menangani permasalahan-permasalahan terkait data dalam bidang sains data (Muller et al. 2019). Profesi baru pun turut ditemukan karena adanya kebutuhan terkait proses sains data misalnya seperti analis data, saintis data, *data engineer*, *analisis business intelligence*, *machine learning engineer*, dll.

Komunitas-komunitas dan lembaga akademi sains data di Indonesia juga semakin menjamur. Selain itu, di bidang pendidikan, Perguruan Tinggi mulai beradaptasi dalam menyesuaikan permintaan lulusan profesi saintis data. Sayangnya, tantangan yang ditemukan yaitu Indonesia dapat dibidang harus membangun lingkungan data yang lebih baik untuk menghimpun dataset yang diperlukan. Perbandingan gaji profesi saintis data dengan saintis informasi yaitu distribusi saintis data bisa lebih dari 20k sementara distribusi saintis informasi kurang dari 15k (Wu and Liu 2019).

Exploratory Data Analysis (EDA) digunakan dalam memahami dataset yang telah terkumpul atau dari berbagai sumber data (Arora, Rajput, and Changotra 2021) (Tummers et al. 2020). Analisis tersebut kemudian dapat memberikan visualisasi deskriptif yang telah terekam oleh data (Lester, Cho, and Lochmiller 2020) (Dsouza and Senthil Velan 2020). EDA dapat dilakukan dengan berbagai tools yang telah tersedia.

Pada artikel ini, peneliti akan menganalisis secara visual pekerjaan sains data di Australia dengan menggunakan pendekatan *Exploratory Data Analysis*. Diharapkan bahwa artikel ini dapat memberikan gambaran mengenai lapangan kerja terkait sains data.

II. METODE PENELITIAN

2.1. Identifikasi Permasalahan

Permasalahan posisi lowongan kerja terkait Data Scientist diambil dikarenakan populernya pekerjaan terkait ilmuwan data. Eksplorasi dibutuhkan untuk memberikan gambaran mengenai pekerjaan terkait ilmuwan data.

2.2. Pengumpulan Data

Dataset Pekerjaan Sains Data di Australia didapatkan dari Kaggle. Kaggle merupakan situs kumpulan dataset. Dataset tersebut digali dari situs *glassdoor* yaitu situs yang mempertemukan pencari kerja dengan lapangan kerja.

2.3. Pembersihan Data

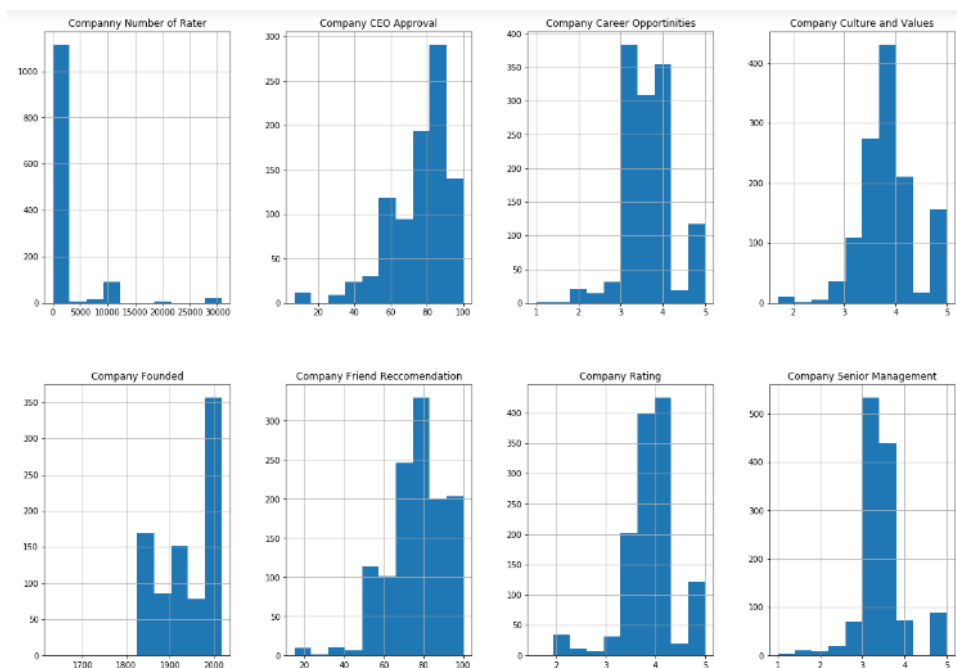
Pembersihan data dilakukan agar data dapat dengan mudah dieksplorasi. Salah satu feature yang dibersihkan yaitu terkait posisi lowongan kerja, Beberapa lowongan tidak memiliki kesamaan nama posisi sehingga yang berkaitan dengan Research, diberi label Researcher, yang berkaitan dengan data arsitektur, diberi label Data Engineer, dll.

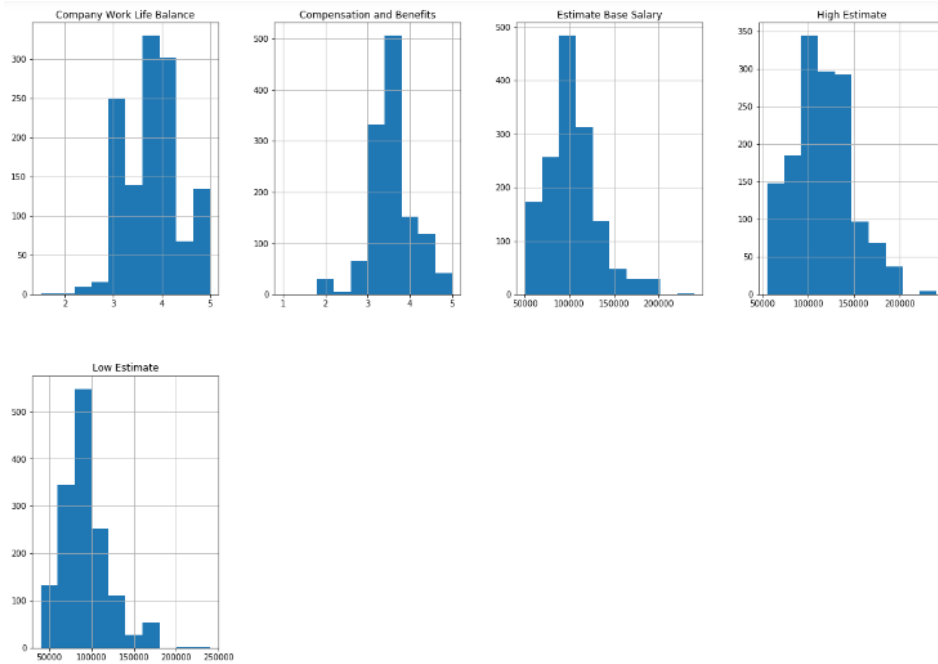
2.4. Eksplorasi Data

Pada tahap ini, eksplorasi dapat dilakukan agar memberikan gambaran mengenai posisi lowongan kerja terkait ilmuwan data.

III. HASIL DAN PEMBAHASAN

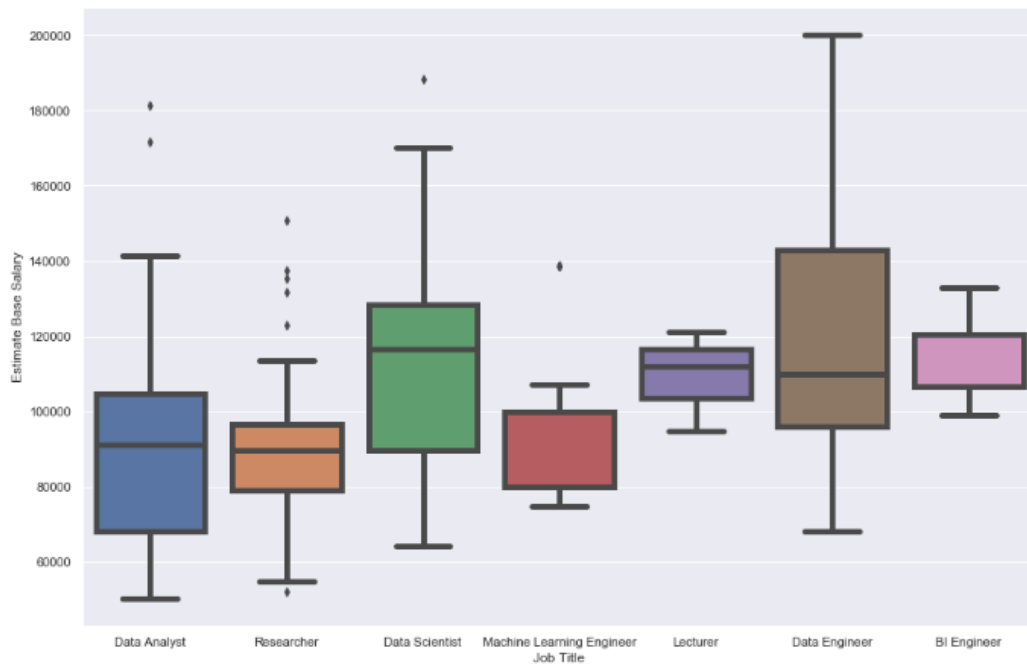
Data yang digunakan dari dataset yaitu *Company Number of Rater*, *Company CEO Approval*, *Company Career Opportunities*, *Company Culture and Values*, *Company Founden*, *Company Friend Recommendation*, *Company Rating*, *Company Senior Management*, *Company Work Life Balance*, *Compensation and Benefits*, *Estimate Base Salary*, *High Estimate*, dan *Low Estimate*. Simpulan distribusi dari variabel tersebut dapat dilihat pada gambar 1



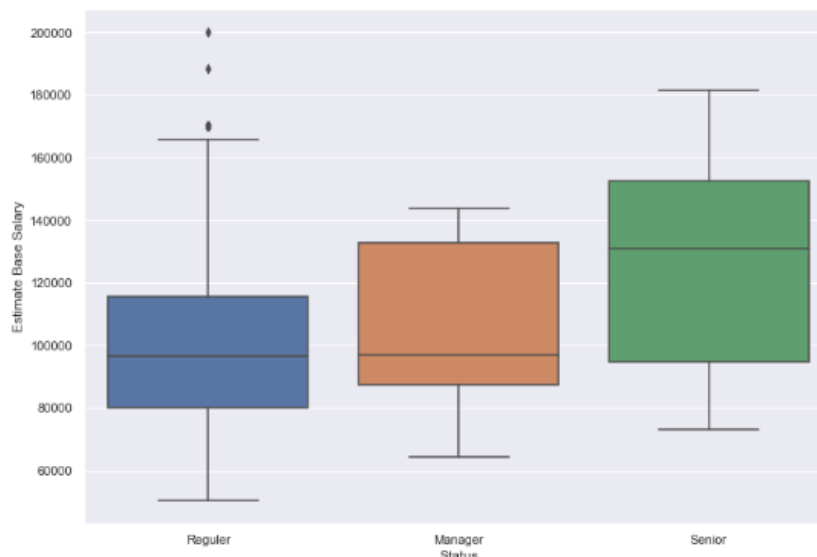


Dari gambar didapatkan bahwa distribusi pengguna yang memberi rating ke perusahaan 0-5000. Persetujuan dari CEO Perusahaan 80-100. Peluang kerja dan nilai dan budaya perusahaan ratingnya 3-4.

Berikutnya terdapat boxplot dari posisi lowongan kerja dengan gaji. Posisi lowongan kerja yang terkait ilmuwan data di Australia yaitu *Data Analyst*, *Researcher*, *Data Scientist*, *Machine Learning Engineer*, *Lecturer*, *Data Engineer*, dan *BI Engineer*. Pada gambar tersebut terdapat outliers di boxplot *Data Analyst*, *Researcher*, *Data Scientist*, *Machine Learning Engineer*. *Data Engineer* memiliki gaji yang paling maksimum, sedangkan *Data Analyst* yang paling minimum. Tiga posisi lowongan kerja dengan nilai median gaji paling tinggi adalah *Data Scientist*, *Lecturer*, dan *Data Engineer*.



Pada situs *Glassdoor* terdapat tingkatan status yang berbeda dari posisi lowongan kerja, *regular* atau *staff* biasa, kemudian *Senior / Lead*, dan ada juga *Manager*. Namun ternyata, pada gaji *regular* yang outlier bahkan bisa melebihi gaji *manager* dan *senior*.



Dari sebaran posisi lowongan kerja dengan tingkatan kerja kita dapat melihat bahwa posisi *regular* menempati jumlah yang paling banyak dicari. Disusul dengan posisi *senior*. Posisi lowongan kerja pada tingkat *regular* paling banyak berturut-turut yaitu *Data Scientist*, *Data Analyst*, dan *Researcher*.

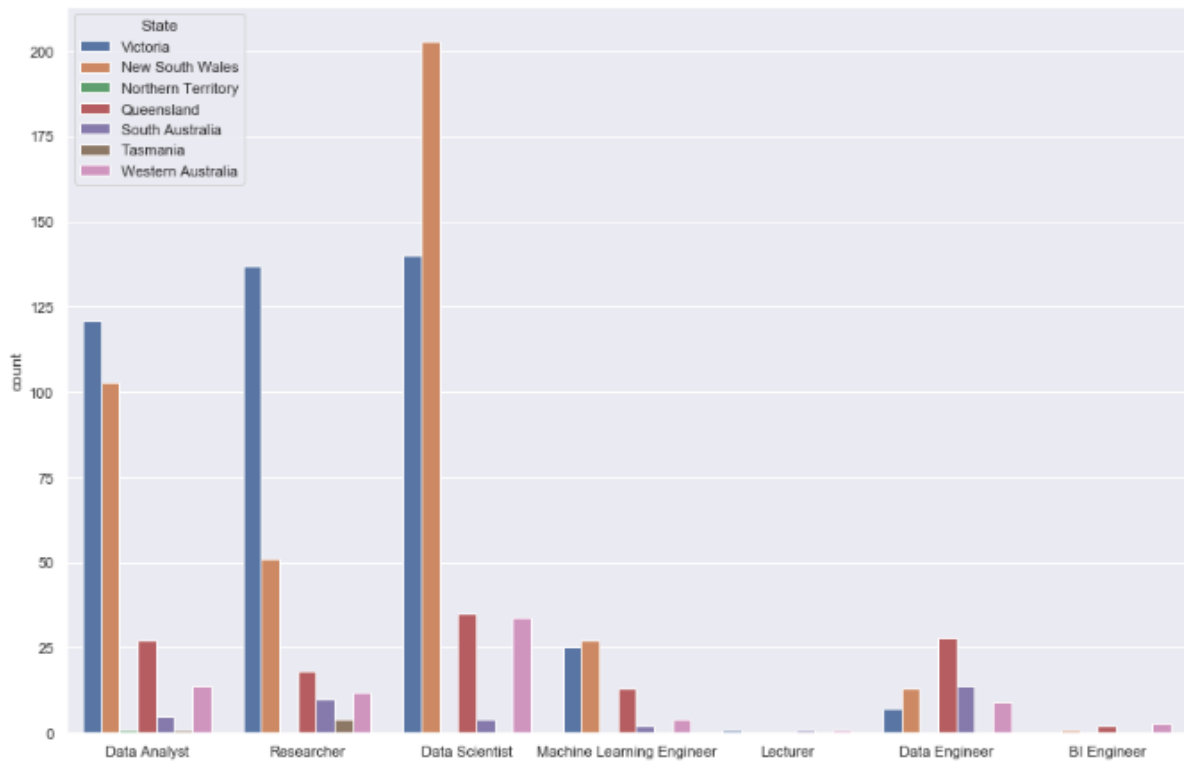


	Status	Manager	Reguler	Senior
Job Title				
BI Engineer		1	4	1
Data Analyst		0	268	6
Data Engineer		2	89	0
Data Scientist		10	399	7
Lecturer		0	2	1
Machine Learning Engineer		0	71	0
Researcher		1	227	4

Lokasi perusahaan juga dapat memberi gambaran mengenai sebaran kebutuhan posisi lowongan kerja. Northern Territory dan Tasmania menempati urutan terakhir sebaran pada lowongan kerja ilmuwan data di Australia. Posisi lowongan kerja paling banyak ditawarkan di New South Wales dan Victoria

State	New South Wales	Northern Territory	Queensland	South Australia	Tasmania	Victoria	Western Australia
Job Title							
BI Engineer	1	0	2	0	0	0	3
Data Analyst	103	1	27	5	1	121	14
Data Engineer	13	0	28	14	0	7	9
Data Scientist	203	0	35	4	0	140	34
Lecturer	0	0	0	1	0	1	1
Machine Learning Engineer	27	0	13	2	0	25	4
Researcher	51	0	18	10	4	137	12

Posisi lowongan kerja sebagai Data Scientist merupakan yang paling banyak dicari di seluruh lokasi di Australia. Yang kedua yaitu Data Analyst, dan yang ketiga yaitu Researcher. Sementara untuk Lecturer memiliki lowongan kerja yang sangat sedikit.



Judul artikel seharusnya dapat mencerminkan inti dari isi tulisan dan ditulis secara ringkas dan lugas.

IV. KESIMPULAN

Ilmuwan data saat ini merupakan karir yang populer dan banyak dibutuhkan oleh industri. Hal ini dibuktikan dengan banyaknya posisi lowongan kerja terkait ilmuwan data yang ditawarkan. Beberapa profesi tersebut antara lain *Data Analyst*, *Researcher*, *Data Scientist*, *Machine Learning Engineer*, *Lecturer*, *Data Engineer*, dan *BI Engineer*. Pada dataset dibahas mengenai gaji dan lokasi terkait profesi ilmuwan data di Australia. Harapan ke depannya di Indonesia juga terdapat dataset tersebut dan juga variabel yang lebih banyak sehingga lebih memberikan *insight* terhadap profesi ilmuwan data di Indonesia.

REFERENSI

1. Arora, Amarpreet Singh, Himadri Rajput, and Rahil Changotra. 2021. “Current Perspective of COVID-19 Spread across South Korea: Exploratory Data Analysis and Containment of the Pandemic.” *Environment, Development and Sustainability* 23 (5): 6553–63. <https://doi.org/10.1007/s10668-020-00883-y>.
2. Dsouza, Joanita, and S. Senthil Velan. 2020. “Using Exploratory Data Analysis for Generating Inferences on the Correlation of COVID-19 Cases.” *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*, 2–7. <https://doi.org/10.1109/ICCCNT49239.2020.9225621>.
3. Lester, Jessica Nina, Yonjoo Cho, and Chad R. Lochmiller. 2020. “Learning to Do Qualitative Data Analysis: A Starting Point.” *Human Resource Development Review* 19 (1): 94–106. <https://doi.org/10.1177/1534484320903890>.
4. Muller, Michael, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. “How Data Science Workers Work with Data.” *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. <https://doi.org/10.1145/3290605.3300356>.
5. Tummers, J., C. Catal, H. Tobi, B. Tekinerdogan, and G. Leusink. 2020. “Coronaviruses and People with Intellectual Disability: An Exploratory Data Analysis.” *Journal of Intellectual Disability Research* 64 (7):



6. Wu, Dan, and Zijun Liu. 2019. “A Comparative Study on Data Science and Information Science: From the Perspective of Job Market Demands in China.” *iConference 2019 Proceedings*.