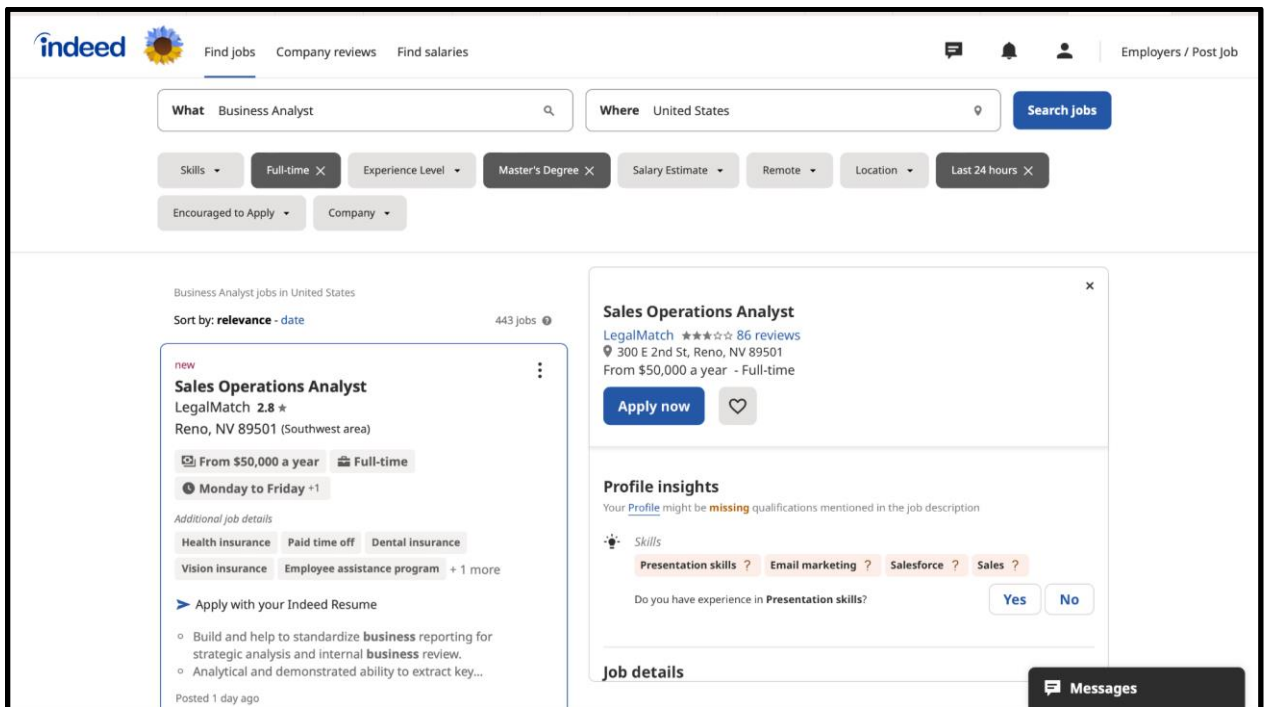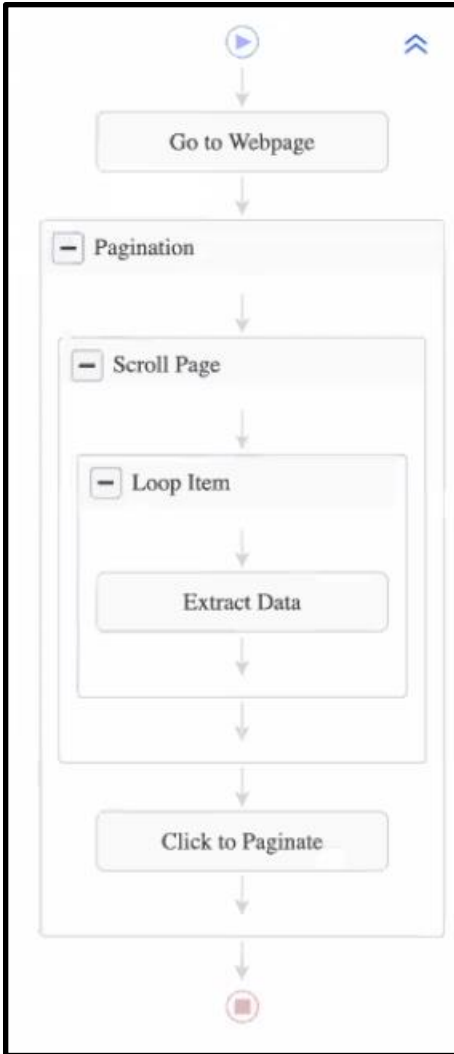# Readme File

I.   Data Collection:

Collected most of the accessible data required for our project, from the following two websites - Indeed & Glassdoor

- Data was filtered for all the full time, Masters level jobs posted 1 day ago across Job Functions of Business Analyst, Data Analyst, Data Engineer & Data Scientist from Indeed

- Inserted the website in Octoparse,APIfy and created a workflow to extract information required to build our database such that all the required data (highlighted in red in below screenshot) is collected in the form of columns and automatically clicks load more jobs button, when it comes to the page end and re-collects the information again till it exhausts all the job listings on the website.

- The web scraped data for each of the job functions - business analyst, data analyst, data scientist and data engineer was extracted to csv files

	Title	Title_URL	Title_URL1	Title2	ratinglink_URL	Number	Company	attribute_snippet	metadata
1	Title	Title_URL	Title_URL1	Title2	ratinglink_URL	Number	Company	attribute_snippet	metadata
2	Takeda Pharmaceutical	https://www.indeed.com/cmp/Ti	https://wv	Senior Manager, Digital Process A	https://www.indeed.coi	3.7	Remote in	$130,200 - $186,00	Full-time
3	Ollie's Bargain Outlet, Inc	https://www.indeed.com/cmp/O	https://wv	Business Analyst	https://www.indeed.coi	2.9	Harrisburg	Full-time	Full-time
4	Super Bright LEDs		https://wv	Business Analyst			Earth City,	$65,000 - $80,000	Full-time
5	NGL Energy Partners LP	https://www.indeed.com/cmp/N	https://wv	Senior Business Systems Analyst	https://www.indeed.coi	2.9	Tulsa, OK	Full-time	Full-time
6	SEQUENTIAL TECHNOLOGY INTERNATIONAL LLC	https://www.indeed.com/cmp/Si	https://wv	Jr Business Analyst	https://www.indeed.coi	3.3	Remote in	Full-time	Full-time
7	Meridian Soft Info Systems		https://wv	Business Analyst			Chicago, IL	$100,000 - $120,00	Full-time
8	OB Hospitalist Group	https://www.indeed.com/cmp/O	https://wv	Business Analyst	https://www.indeed.coi	3.0	Greenville,	$68,000 - $76,000	Full-time
9	Jada Systems, Inc.		https://wv	Business Analyst			Los Angele	$63,819 - $135,994	Full-time+
10	TekisHub Consulting Services		https://wv	Clinical Business Analyst			Lorain, OH	$65 - $70 an hour	Contract
11	HUSCO International, Inc.	https://www.indeed.com/cmp/H	https://wv	Business Data Analyst	https://www.indeed.coi	3.3	Waukesha	Full-time	Full-time
12	Heritage Operations Group, LLC		https://wv	Business Analyst			Bloomingt		
13	Fiserv, Inc.	https://www.indeed.com/cmp/Fi	https://wv	Business Analyst	https://www.indeed.coi	3.4	Fishers, IN	Full-time	Full-time
14	Input 1	https://www.indeed.com/cmp/Ir	https://wv	Business Systems Analyst	https://www.indeed.coi	3.2	+1 locatioi	$70,000 - $90,000	Full-time
15	Orangetheory - Corporate Headquarters - #0001	https://www.indeed.com/cmp/O	https://wv	Business Intelligence Analyst II	https://www.indeed.coi	3.4	Hybrid ren	Full-time	Full-time
16	iMedia Brands, Inc	https://www.indeed.com/cmp/Ir	https://wv	Business Analyst, Call Center	https://www.indeed.coi	2.6	Bowling Gi	Full-time	Full-time
17	NationsCX	https://www.indeed.com/cmp/N	https://wv	Project Manager	https://www.indeed.coi	2.3	Plantation	From $70,000 a yea	Full-time
18	Talntly	https://www.indeed.com/cmp/Ti	https://wv	PM/BA	https://www.indeed.coi	5.0	Remote	$45 - $60 an hour	Full-time+
19	TEKsystems	https://www.indeed.com/cmp/Ti	https://wv	Business Analyst	https://www.indeed.coi	3.9	East Lansir	Contract	Contract
20	Better Business Bureau Serving the Heart of Texas	https://www.indeed.com/cmp/Bi	https://wv	Business Practice Analyst	https://www.indeed.coi	3.5	Hybrid ren	Full-time	Full-time
21	TRESUME		https://wv	Entry Level Business Analyst/Data			Owings Mi	$26 - $30 an hour	Full-time+

Sheet1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Title	Title_URL	Title1	Label	Company	Company	ratinglink	Number	attribute	metadata	metadata	ialbl	jobsnippe	jobsnippe	jobsn
2	Brockton, M.	https://www	Data Analyst	new	VA Boston H				Full-time	Full-time	Monday to F	Easily apply	Working kno	data	
3	Hybrid remo	https://www	Data Analyst	new	Blue Cross Bl				$55,000 - $6!	Full-time+1	Monday to F	Easily apply	Ability to an:	data	data
4	Remote	https://www	Business Syst	new	EclipseIA	https://www	https://www	2.5	Full-time	Full-time	Monday to F	Easily apply	Develop SQL	data	data
5	Providence, I	https://www	Health Data	new	Rhode Island	https://www	https://www	3.7	Full-time	Full-time		Easily apply	Two+ years r	data	data
6	Hybrid remo	https://www	Business Dat	new	Ascendo Res	https://www	https://www	3.4	$50,000 - $5!	Full-time	Monday to F	Easily apply	1+ years of e	data	
7	Bellevue, W/	https://www	Data Analyst	new	Veridants te				$60,859 - $1:	Full-time	8 hour shift	Easily apply	5-7 years of e	data	Data
8	Salem, OR 97	https://www	Healthcare D	new	Northwest H	https://www	https://www	3.5	$56,376 - $8!	Full-time	Monday to F	Easily apply	Bachelor's d	Data	
9	Largo, FL 337	https://www	Data Analyst	new	Bioderm Inc.	https://www	https://www	3.6	$100,000 - $:	Full-time		Easily apply	Offers insigh	data	data
10	Phoenix, AZ I	https://www	Data Analyst	new	Arizona Vete				$55,000 - $8(	Full-time		Easily apply	Demonstrate	data	data
11	Baltimore, N	https://www	Entry Level D	new	TRESUME				$26 - $30 an	Full-time+1	Monday to F	Easily apply	Includes exp	data	data
12	Remote in Ci	https://www	Clinical Data	new	Vetro Tech Ir				$71,815 - $8!	Full-time	Monday to F	Easily apply	Proven work	data	analy
13	Hybrid remo	https://www	Data Analyst	new	OrboGraph				$80,000 - $1:	Full-time	Monday to F	Easily apply	In addition, t	data	analy
14	Remote	https://www	Data Analyst	new	Heyday Skin				$90,000 - $1(	Full-time		Easily apply	Perform data	data	data
15	Parsippany, I	https://www	Data Analyst	new	DOWC	https://www	https://www	4.0	Full-time	Full-time	Monday to F	Easily apply	Assist in the	data	data
16	Hybrid remo	https://www	Claims and D	new	Empower He				From $84,30	Full-time	Monday to F	Easily apply	Experience v	data	data
17	Chicago, IL+1	https://www	Business Ana	new	GNS STAFFIN				$12.00 - $31.	Full-time	Monday to F	Easily apply	1+ years of P		
18	Remote in Si	https://www	Sr. Product A	new	First America	https://www	https://www	3.8	Full-time	Full-time			Enhance exis	data	data
19	Basking Ridg	https://www	Data Analyst	new	RMANJ Man	https://www	https://www	3.5	Full-time	Full-time		Easily apply	Fulfill ad-hoc	data	
20	Remote in Tr	https://www	Research An:	new	Ipsos North /	https://www	https://www	3.9	Full-time	Full-time		Easily apply	Liaising with	data	data
21	Remote in Ci	https://www	Data Analyst	new	NewAge Prot	https://www	https://www	3.8	Full-time	Full-time		Easily apply	Minimum of	data	analy
22	Fort Bragg, N	https://www	Data Analyst	new	Occam Solut	https://www	https://www	3.9	Full-time	Full-time		Easily apply	Build data sc	data	data

- Collected Firm Ratings for companies from the web scraped data for Rating table in our database from both Indeed and Glassdoor
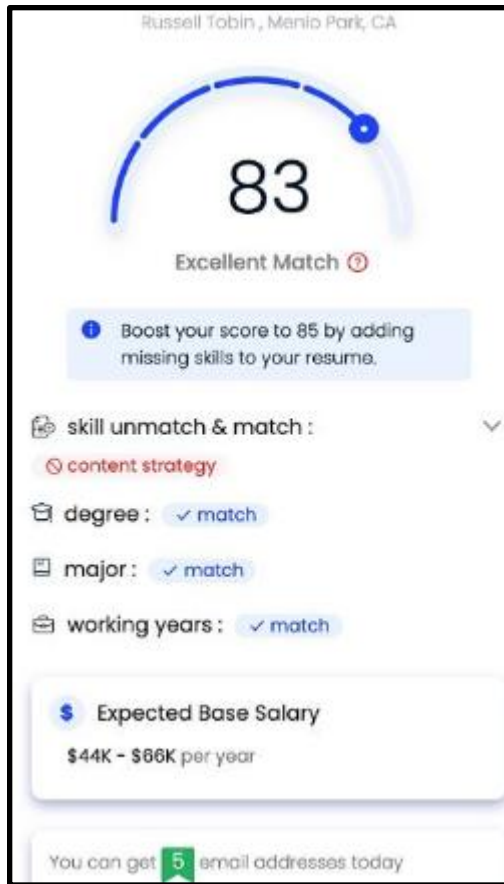
- Indeed
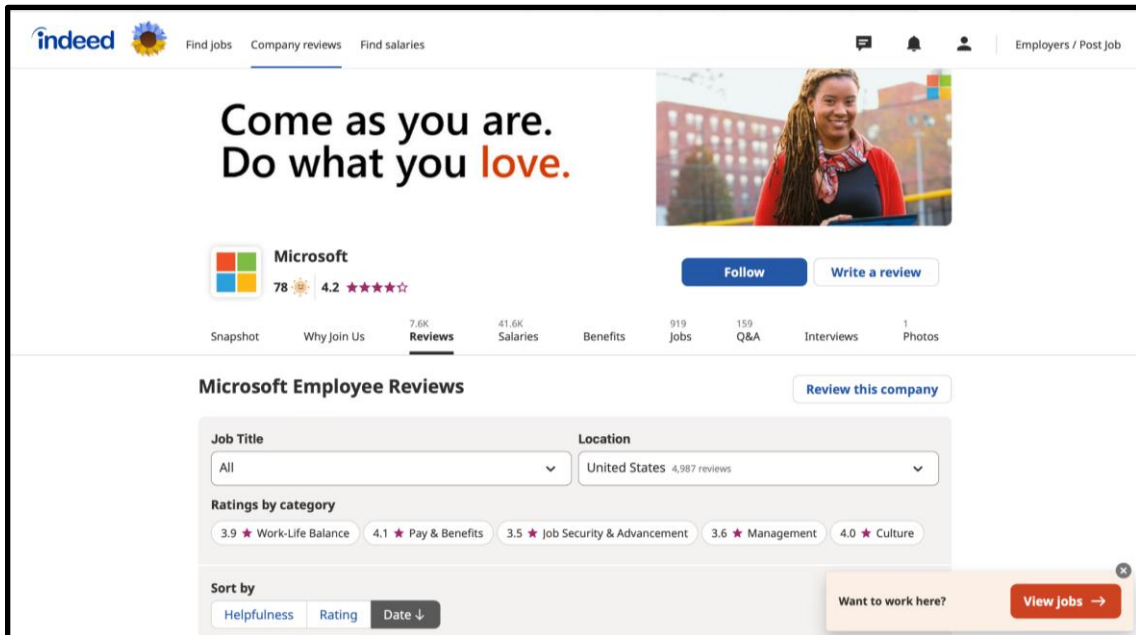
- Glassdoor
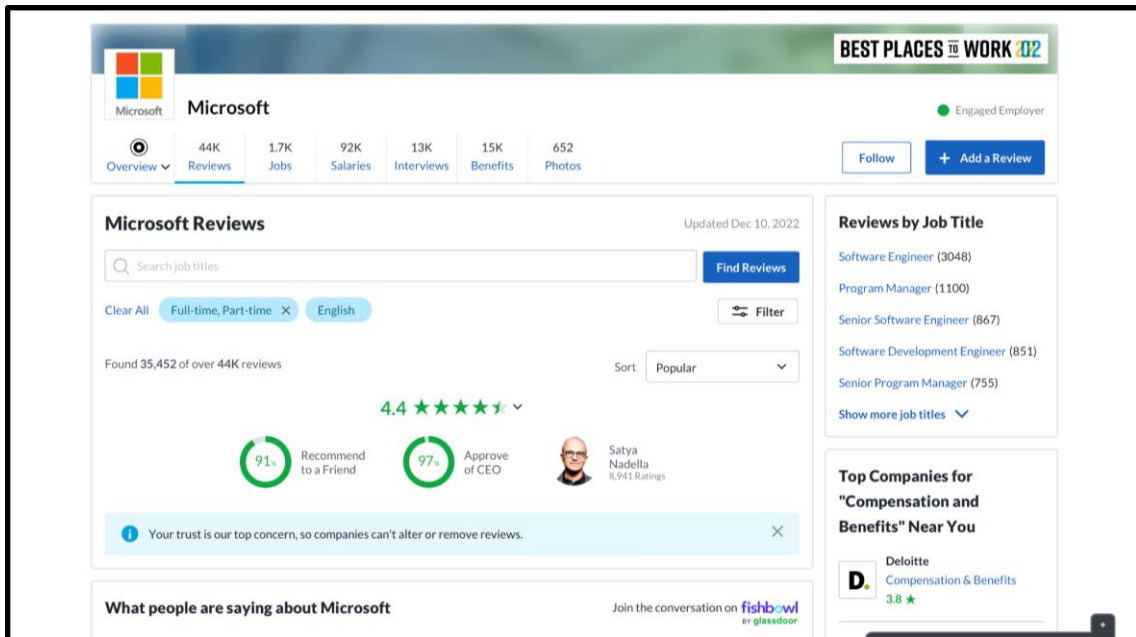
- Collected Resumes and Information from few University of Maryland, CP students and matched their profiles with JobSeer tool to the extracted Job data.
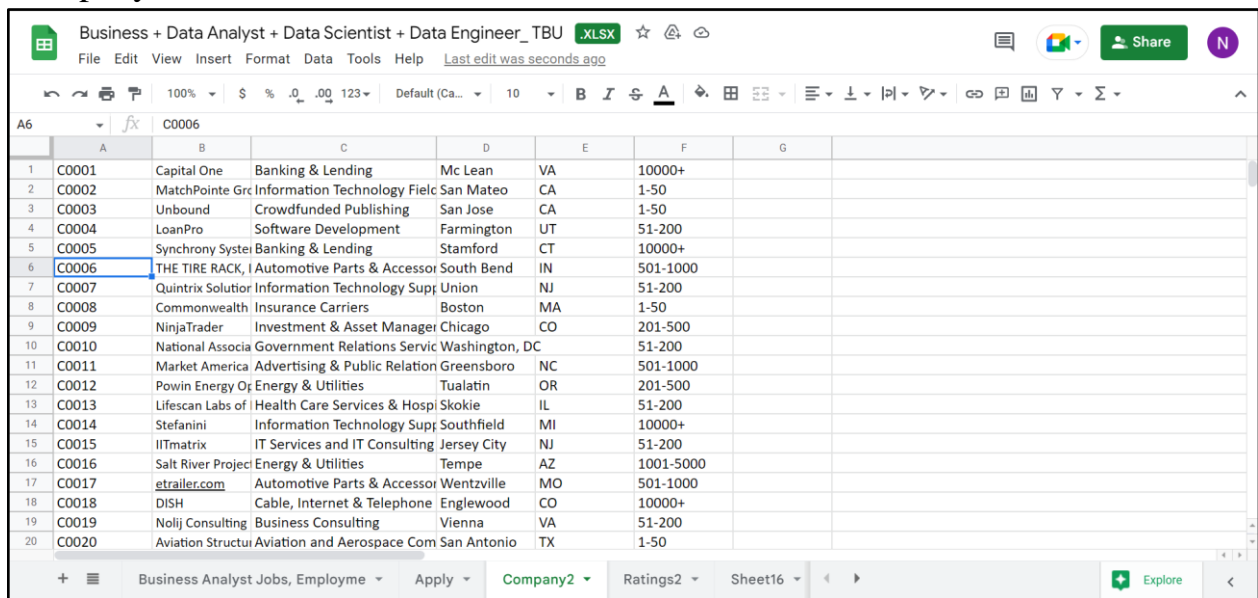
II. Data Pre-processing:

- The raw data extracted in csv format was further processed to remove duplicate values, null values and redundant information using **Remove Duplicates** in Microsoft Excel.
- The job Location was mapped to either Remote or 2 letter State Code.
- The job Salary was converted to per annum from monthly or hourly formats and averages were taken in cases of salary range in order to make the data uniform for analysis purposes.
- Assigned unique Identifiers for each of Job, company, JobSeeker, Rating, ProfileMatch tables etc.
- The data, in its raw form was converted to a more readable format, that could be then entered into database tables. The new finalized data has headings, no null values and no duplicate values.

Post cleaning of the data, the below is the finalized set of data stored in excel.

Example:

Company:

III. REFERENCES
- https://www.indeed.com/
- https://www.glassdoor.com/

IV. TEST METHODS

To test our database, and to be able to view and extract required information, we have used sql queries and tableau visualizations. Below are the associated results.

Using the dashboard created , users can see the data associated with each job function for different business transactions. One such example is shown below

Top 5 Companies by Salary within each Job Function

Job Function	Company Name	Salary
Data Analyst	Deloitte	173,000
	Hertz	163,150
	Verizon	161,000
	ALDI	154,000
	Castellini Company LLC	153,000

Salary (in $ per year): 0K 50K 100K 150K 200K

Job Function
Data Analyst

Companies with Job Openings in Multiple Job Functions

Job Function
Data Analyst

ALDI
Cushman & Wakefield
DISH
FocusKPI Inc.
PCS Global Tech
U.S. Bank
Wolverine Trading

Percentage of Jobs Requiring Skills by Job Function

Skill Name	Job Function Data Analyst
Apache Hadoop	32.65%
Apache Spark	26.53%
Communication	8.16%
Crticial Thinkin	30.61%
Java	26.53%
Machine Learning	63.27%
MATLAB	2.04%
Microsoft AWS	24.49%
Microsoft Azure	22.45%
Microsoft Excel	14.29%
Power BI	20.41%
Powerpoint	4.08%
Presentation Sk..	6.12%
Python	77.55%
Qlik Sense	16.33%
R	61.22%
SAP	18.37%
SAS	6.12%
SnowFlake	8.16%
SQL	51.02%
Statistics	63.27%
Tableau	83.67%