The Perfect Analyst

Analysis of Data Oriented Jobs

This proposal contains a detailed description of the plan and design of database to use Data Oriented Job Application Tool

Created By:
Madathil Geetanjali Menon
Sri Sai Alekya Ghanta
Aishwarya Sadagopan
Nishanthi Ravichandran

ABOUT

Mission Statement

The project's mission is to address certain common queries of Masters's students seeking jobs, particularly those applying for a full-time Data Oriented / Analyst role in the United States, via Indeed.com.

Mission Objectives

The project would focus on the following main deliverables:
Creating the relevant Database, consisting of the tables and views for the Entities (JobSeeker, Job, Company, Rating, Skill, ProfileMatch)

- Constructing an ER diagram and the Relational Schema, that would capture all the stated nuances.
- To build SQL queries to be able to extract data as per the Client's requirement and present the same. These include the following:
    - From a Job Seeker's point of view:
        - Top companies offering high salaries across each job function
        - Identify top skills required for each of the data oriented roles
        - Identify best-rated firms as per the reviews from Glassdoor and Indeed
        - Determine companies offering remote vs onsite opportunities
        - Companies with openings in more than one job function
        - To identify job seeker's profile match rate with different jobs
    - From a Company's / Organization's point of view:
        - Understand reviews by current employees and leverage them for the company's development
        - Identify top candidates for open positions, based on the required skills

BACKGROUND

Goals:

- Address common queries of Masters' students seeking full-time job opportunities in the Data Oriented / Analyst job functions across the United States, via Indeed.com

End Users:

- Graduate Students, Organizations' Talent Acquisition Team are our target users

Data Sources:

The technique used to obtain data: Web Scraping
Data has been scraped from various sources such as:

- Indeed
- Glassdoor

The tools used to scrape data:

- Octoparse
- APIFY

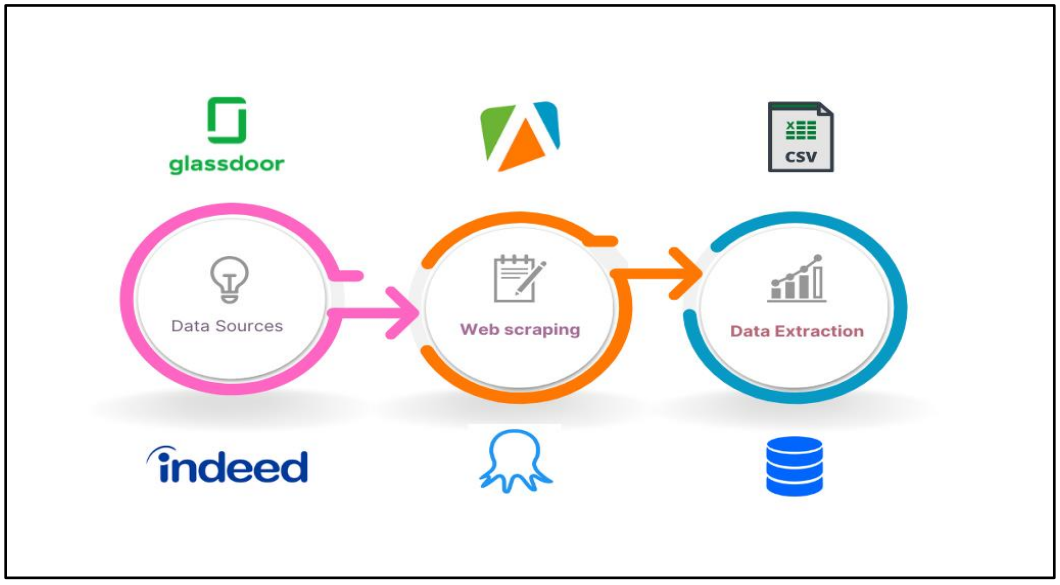


Figure 1: Representation of Data Sources

BUSINESS PROCESS AND TRANSACTIONS

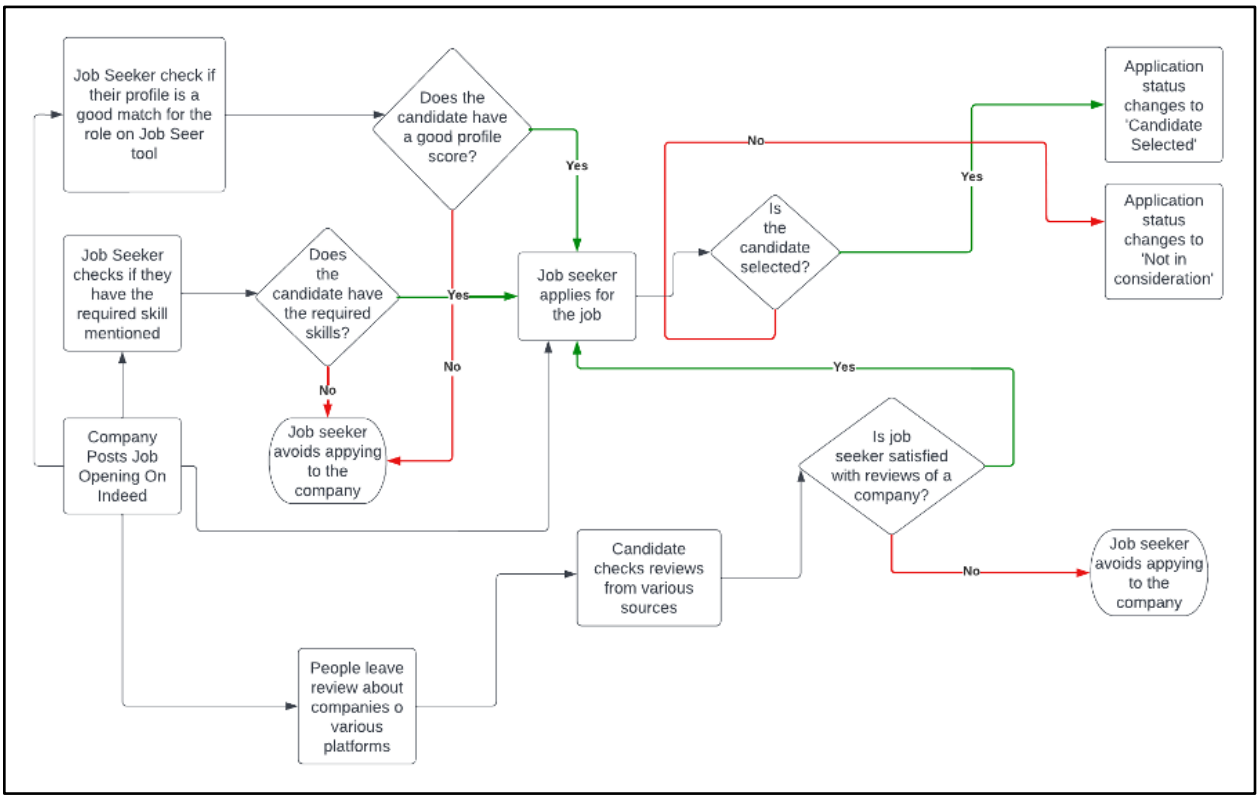


Figure 2: Business process

DATABASE ANALYSIS

ER Schema

Entities, Attributes and Primary Keys

JobSeeker(**jbsId**, jbsName, -jbsFirstName,-jbsLastName, jbsBirthDate, jbsEmailId, jbsPhnNum, jbsAddress, -jbsCity, -jbsState, -jbsZipCode, jbsEducation[1..3], jbsPrevExperience[1..5])

Company(**compId,** compName, compDesc, compLocation, -compCity, -compState, compSize)

Job(**<u>jobId</u>**, jobFunction, jobDesc, jobSalary, jobType, jobLocation, =avgSalaryByFunction)

Rating(**<u>ratId</u>**, ratCultureValue, ratWorkLifeBalance, ratSeniorManagement, ratCompBenefits, ratCareerOpportunities, ratOverallRating, =avgRating, ratSiteSource)

Skill(**<u>sklId</u>**, sklName)

ProfileMatch(**<u>pmId</u>**, pmScore, pmCategory, pmSkillMatch, pmNumOfAlumni)

<u>Relationships, Attributes, Degrees, Participating Entities and Constraints</u>

Require: Ternary Relationship
1 JobSeeker and 1 Skill to 0 or many Jobs
1 Skill and 1 Job to 0 or many JobSeekers
1 JobSeeker and 1 Job to 1 or many Skills

Has: Binary Relationships
1 JobSeeker to 1 or many Skills
1 Skill to 0 or many JobSeekers

Apply(appDate, appStatus**):** Ternary Relationship
1 Job and 1 Company to 0 or many JobSeekers
1 Company and 1 JobSeeker to 0 or many Jobs
1 JobSeeker and 1 Job to 1 Company only

Review: Binary Relationship
1 Company to 0 or many Ratings
1 Rating to 1 Company

Posts: Binary Relationship
1 Company to 1 or many Jobs
1 Job to 1 Company

Check: Ternary Relationship
1 JobSeeker and 1 Job to 0 or 1 ProfileMatch
1 ProfileMatch and 1 Job to 1 JobSeeker
1 JobSeeker and 1 ProfileMatch to 1 Job

ER diagram

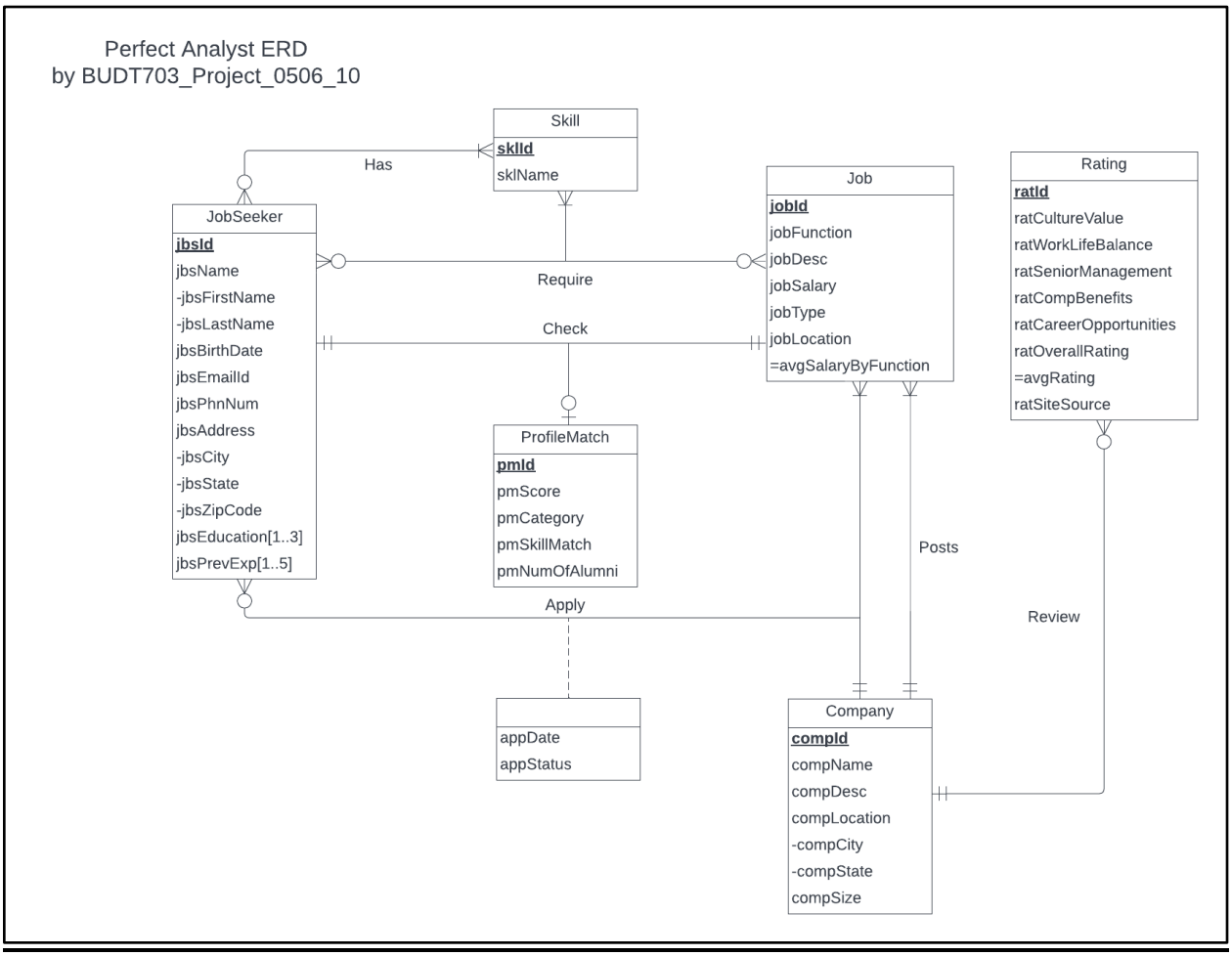


Figure 3: ER Diagram

Relational Schema:

JobSeeker (**jbsId**, jbsFirstName, jbsLastName, jbsBirthDate, jbsEmailId, jbsPhnNum, jbsCity, jbsState, jbsZipCode)

Job Seeker Education (***jbsId***, **jbsEducation**)

Job Seeker Previous Experience (***jbsId***, **jbsPrevExperience**)

Company (**compId**, compName, compDesc, compCity, compState, compSize)

Job (**jobId**, jobFunction, jobDesc, jobSalary, jobType, jobLocation, *compId*)

Rating (**ratId**, ratCultureValue, ratWorkLifeBalance, ratSeniorManagement, ratCompBenefits, ratCareerOpportunities, ratOverallRating, ratSiteSource, *compId*)

Skill (**sklId**, sklName)

ProfileMatch (**pmId**, pmScore, pmCategory, pmSkillMatch, pmNumOfAlumni)

Has (***jbsId***, ***sklId***)

Require (***jbsId***, ***jobId***, ***sklId***)

Check (***pmId***, *jbsId*, *jobId*)

Apply (***jbsId***, ***jobId***, *compId*, appDate, appStatus)

Business rules:

[R1]: For a JobSeeker having certain Skills, the JobSeeker information cannot be deleted from the database.
[R2]: For a JobSeeker having certain Skills, if the JobSeeker information is updated, then the corresponding Skill information must be updated.
[R3]: For a JobSeeker having certain Skills, the Skill information cannot be deleted as the JobSeeker already has acquired those Skills.
[R4]: For a JobSeeker having certain Skills, if the Skill information is updated, then the corresponding JobSeeker information must be updated.
[R5]: When a JobSeeker seeks a Job that requires certain Skills, the JobSeeker information cannot be deleted from the database.
[R6]: When a JobSeeker seeks a Job that requires certain Skills, if the JobSeeker information is updated, then the corresponding Job and Skill information should be updated.
[R7]: For a particular Job that requires certain Skills from a JobSeeker, if the job information is deleted, then the corresponding JobSeeker and Skill information should be deleted.
[R8]: For a particular Job that requires certain Skills from a JobSeeker, if the Job information is updated, then the corresponding JobSeeker and Skill information should be updated.
[R9]: For a Skill that is required by a Job from a JobSeeker, if the Skill information is deleted, then the corresponding JobSeeker and Job information should be deleted.
[R10]: For a Skill that is required by a Job from a JobSeeker, if the Skill information is updated, then the corresponding JobSeeker and Job information should be updated.
[R11]: For a ProfileMatch being checked by a JobSeeker for a particular Job, if the ProfileMatch information is deleted, then the corresponding Job and JobSeeker information should be deleted.

[R12]: For a ProfileMatch being checked by a JobSeeker for a particular Job, if the ProfileMatch information is updated, then the corresponding Job and JobSeeker information should be updated.
[R13]: For a JobSeeker checking ProfileMatch for a certain job, the JobSeeker information cannot be deleted.
[R14]: For a JobSeeker checking ProfileMatch for a certain Job, if the JobSeeker information is updated, then the corresponding Job and ProfileMatch information should be updated.
[R15]: If the Job information for a particular Job is deleted, then the corresponding details of the ProfileMatch and JobSeeker should also be deleted.
[R16]: If the Job information for a particular Job is updated, then the corresponding details of the ProfileMatch and the JobSeeker should also be updated.
[R17]: For a particular JobSeeker applying to a Job in a Company, then the JobSeeker information cannot be deleted.
[R18]: For a particular JobSeeker applying to a Job in a Company, if the JobSeeker information is updated, then the corresponding details of the Job, Company and application should be updated.
[R19]:If the Job information of a particular Job is deleted, then the corresponding details of the JobSeeker, Company and application should also be deleted.
[R20]:If the Job information for a particular Job is updated, then the corresponding details of the JobSeeker, Company and application should also be updated.
[R21]: If the Company information of a particular Company is deleted, then the corresponding details of the Job, JobSeeker and application should also be deleted.
[R22]: If the Company information of a particular Company is updated, then the corresponding details of the Job, JobSeeker and application should also be updated.
[R23]: If the Company information of a particular Company is deleted, then the corresponding details of the Rating should also be deleted.
[R24]: If the Company information of a particular Company is updated, then the corresponding details of the Rating should also be updated.
[R25]: For a JobSeeker who has education listed, the JobSeeker information cannot be deleted.
[R26]: For a JobSeeker who has education listed, if the JobSeeker information is updated then the corresponding JobSeeker education information should also be updated.
[R27]: For a JobSeeker who has previous experience listed, the JobSeeker information cannot be deleted.
[R28]: For a JobSeeker who has previous experience listed, if the JobSeeker information is updated then the corresponding JobSeeker previous experience information should also be updated.
[R29]: For a Company that has a Job listed, the Job details cannot be deleted from the database.
[R30]: For a Company that has a Job listed, the Job details cannot be updated on the database.

Referential Integrities:

Relation	Foreign Key	Base Relation	Primary Key	Business Rule	Constraint: ON DELETE	Business Rule	Constraint: ON UPDATE
Has	jbsId	JobSeeker	jbsId	R1	NO ACTION	R2	CASCADE
Has	sklId	Skill	sklId	R3	NO ACTION	R4	CASCADE
Require	jbsId	JobSeeker	jbsId	R5	NO ACTION	R6	CASCADE
Require	jobId	Job	jobId	R7	CASCADE	R8	CASCADE
Require	sklId	Skill	sklId	R9	CASCADE	R10	CASCADE
Check	pmId	ProfileMatch	pmId	R11	CASCADE	R12	CASCADE
Check	jbsId	JobSeeker	jbsId	R13	NO ACTION	R14	CASCADE
Check	jobId	Job	jobId	R15	CASCADE	R16	CASCADE
Apply	jbsId	JobSeeker	jbsId	R17	NO ACTION	R18	CASCADE
Apply	jobId	Job	jobId	R19	CASCADE	R20	CASCADE
Apply	compId	Company	compId	R21	CASCADE	R22	CASCADE
Rating	compId	Company	compId	R23	CASCADE	R24	CASCADE
Job Seeker Education	jbsId	JobSeeker	jbsId	R25	NO ACTION	R26	CASCADE
Job Seeker Previous Experience	jbsId	JobSeeker	jbsId	R27	NO ACTION	R28	CASCADE
Job	compId	Company	compId	R29	NO ACTION	R30	NO ACTION

IMPLEMENTING DATABASE

After the development of the database, 6 business transactions have been configured on SQL and Tableau as examples.

BUSINESS TRANSACTION 1: ***What are the top 5 companies by salary within each job function?***

SQL CODE:

```
SELECT a.jobFunction AS 'Job Function', a.compName AS 'Company Name' ,a.[jobSalary (in $ per year)] AS 'Salary (in $ per year)'
FROM (SELECT j.jobFunction, c.compName ,j.[jobSalary (in $ per year)],
            ROW_NUMBER() OVER (PARTITION BY j.jobFunction ORDER BY j.[jobSalary (in $ per year)] DESC) as 'Rnk'
      FROM [PerfectAnalyst.Job] j , [PerfectAnalyst.Company] c
      WHERE j.compId = c.compId
     ) a WHERE Rnk<= 5
```

Figure 4: SQL Code for Business Transaction 1

SQL OUTPUT:

Results | Messages

	Job Function	Company Name	Salary (in $ per year)
1	Business Analyst	Home Depot / THD	168000
2	Business Analyst	DHL | Supply Chain	162100
3	Business Analyst	Takeda Pharmaceutical	158100
4	Business Analyst	IBM	155000
5	Business Analyst	The Baton Rouge Clinic, AMC	151000
6	Data Analyst	Deloitte	173000
7	Data Analyst	Hertz	163150
8	Data Analyst	Verizon	161000

nsmith.umd.edu,97... | AD\aishsada (107) | BUDT703_Project_0506_10 | 00:00:00 | 20 rows

Figure 5: SQL Output for Business Transaction 1

TABLEAU:

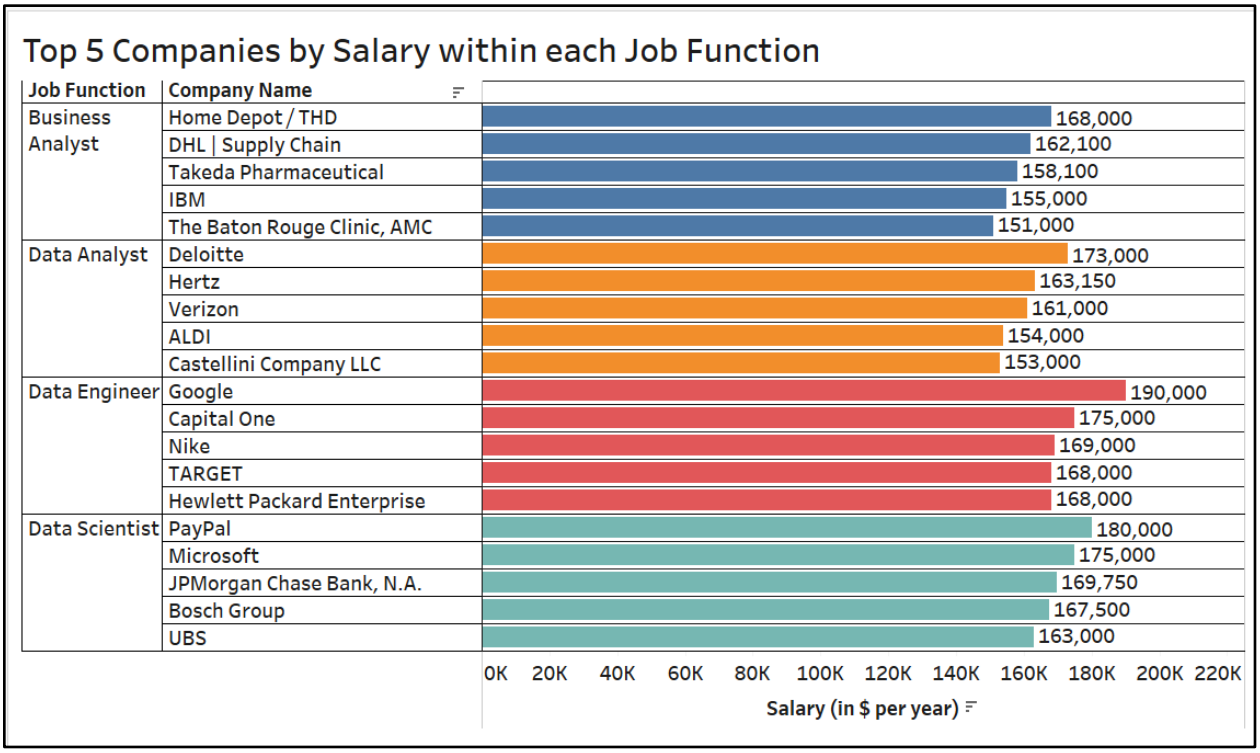


Figure 6: Tableau Output for Business Transaction 1

BUSINESS TRANSACTION 2:

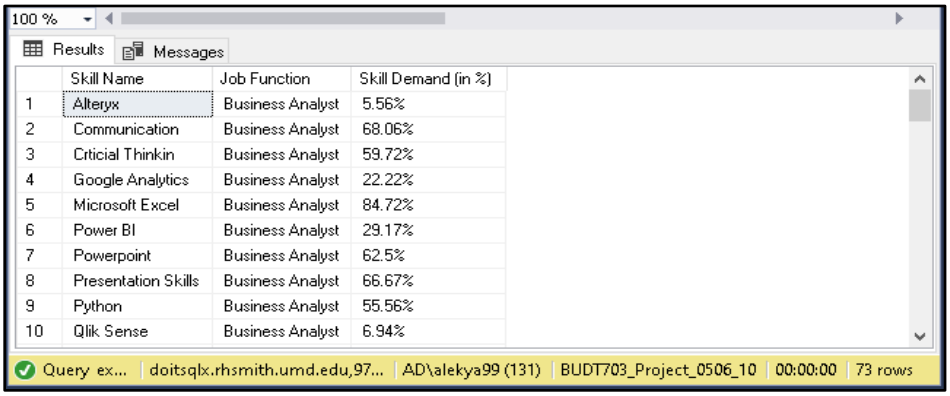
What are the percentage of different skills required in each job function?

SQL CODE:

```
SELECT a.sklName AS 'Skill Name', a.jobFunction AS 'Job Function',
       CONCAT(CAST(ROUND((CAST(a.[Job Count by Skill & Function] AS FLOAT)/CAST(a.[Job Count By Function] AS FLOAT)) * 100,2) AS VARCHAR),'%') AS 'Skill Demand (in %)'
FROM (SELECT DISTINCT s.sklName,j.jobFunction,
                      COUNT(j.jobId) OVER(PARTITION BY s.sklId, j.jobFunction) AS 'Job Count by Skill & Function',
                      a.[Job Count By Function]
      FROM [PerfectAnalyst.Job] j, [PerfectAnalyst.Skill] s, [PerfectAnalyst.Require] r,
           (SELECT j.jobFunction ,COUNT(j.jobId)as 'Job Count By Function'
            FROM [PerfectAnalyst.Job] j
            GROUP BY j.jobFunction) a
      WHERE j.jobid = r.jobid and s.sklId = r.sklId and a.jobFunction = j.jobFunction
      ) a
ORDER BY a.jobFunction , a.sklName
```

Figure 7: SQL Code for Business Transaction 2

SQL OUTPUT:



100 %

Results | Messages

	Skill Name	Job Function	Skill Demand (in %)
1	Alteryx	Business Analyst	5.56%
2	Communication	Business Analyst	68.06%
3	Crticial Thinkin	Business Analyst	59.72%
4	Google Analytics	Business Analyst	22.22%
5	Microsoft Excel	Business Analyst	84.72%
6	Power BI	Business Analyst	29.17%
7	Powerpoint	Business Analyst	62.5%
8	Presentation Skills	Business Analyst	66.67%
9	Python	Business Analyst	55.56%
10	Qlik Sense	Business Analyst	6.94%

Query ex... | doitsqlx.rhsmith.umd.edu,97... | AD\alekya99 (131) | BUDT703_Project_0506_10 | 00:00:00 | 73 rows

Figure 8: SQL Output for Business Transaction 2

TABLEAU:

	Job Function			
Skill Name	Business Analyst	Data Analyst	Data Engineer	Data Scientist
Alteryx	5.56%			
Apache Hadoop		32.65%	56.86%	40.38%
Apache Spark		26.53%	88.24%	42.31%
Communication	68.06%	8.16%		5.77%
Crticial Thinkin	59.72%	30.61%	17.65%	19.23%
Google Analytics	22.22%			
Java		26.53%	41.18%	46.15%
Javascript			31.37%	38.46%
Machine Learning		63.27%	23.53%	86.54%
MATLAB		2.04%	25.49%	
Microsoft AWS		24.49%	54.90%	30.77%
Microsoft Azure		22.45%	60.78%	32.69%
Microsoft Excel	84.72%	14.29%		
Power BI	29.17%	20.41%	29.41%	5.77%
Powerpoint	62.50%	4.08%	5.88%	
Presentation Skills	66.67%	6.12%		
Python	55.56%	77.55%	52.94%	76.92%
Qlik Sense	6.94%	16.33%		
R	2.78%	61.22%	47.06%	73.08%
SAP		18.37%	23.53%	7.69%
SAS		6.12%	11.76%	21.15%
SnowFlake		8.16%	70.59%	25.00%
SQL	76.39%	51.02%	78.43%	57.69%
Statistics	20.83%	63.27%	29.41%	92.31%
Tableau	45.83%	83.67%	33.33%	9.62%

Figure 9: Tableau Output for Business Transaction 2

BUSINESS TRANSACTION 3: ***What are top 10 overall best-rated firms as per reviews from Glassdoor and Indeed?***

SQL CODE:

```
SELECT TOP 10 c.compName AS 'Company Name', FORMAT(AVG(r.ratOverallRating),'N2') AS 'Overall Average',
                                            FORMAT(AVG(r.ratCareerOpportunities),'N2') AS 'Career Opportunties Average',
                                            FORMAT(AVG(r.ratCompBenefits),'N2') AS 'Compensation Benefits Average',
                                            FORMAT(AVG(r.ratCultureValue),'N2') AS 'Culture & Value Average' ,
                                            FORMAT(AVG(r.ratSeniorManagement),'N2') AS 'Senior Management Average',
                                            FORMAT(AVG(r.ratWorkLifeBalance),'N2') AS 'Work Life Balance Average'
FROM [PerfectAnalyst.Company] c, [PerfectAnalyst.Rating] r
WHERE c.compId = r.compId
GROUP BY c.compName
ORDER BY AVG(r.ratOverallRating) DESC, c.compName DESC
```

Figure 10: SQL Code for Business Transaction 3

SQL OUTPUT:

Results | Messages

	Company Name	Overall Average	Career Opportunties Average	Compensation Benefits Average	Culture & Value Average	Senior Management Average	Work Life Balance Average
1	Hertz	5.00	5.00	4.95	5.00	4.95	5.00
2	Capital One	5.00	4.95	5.00	5.00	5.00	5.00
3	TARGET	4.95	5.00	5.00	4.95	5.00	4.95
4	JPMorgan Chase Bank, N.A.	4.95	5.00	5.00	4.95	5.00	5.00
5	Deloitte	4.95	5.00	4.95	4.95	5.00	5.00
6	ALDI	4.95	4.95	4.95	5.00	4.95	4.95
7	Jada Systems, Inc.	4.90	4.90	4.90	4.70	4.90	4.60
8	Cascade Energy Inc	4.80	4.85	4.70	4.85	4.15	4.40
9	Microsoft	4.65	4.75	4.65	4.60	4.65	4.55
10	Google	4.60	4.60	4.20	4.45	4.40	4.35

Query executed successfully. | doitsqlx.rhsmith.umd.edu,97... | AD\aishsada (84) | BUDT703_Project_0506_10 | 00:00:00 | 10 rows

Figure 11: SQL Output for Business Transaction 3

TABLEAU:

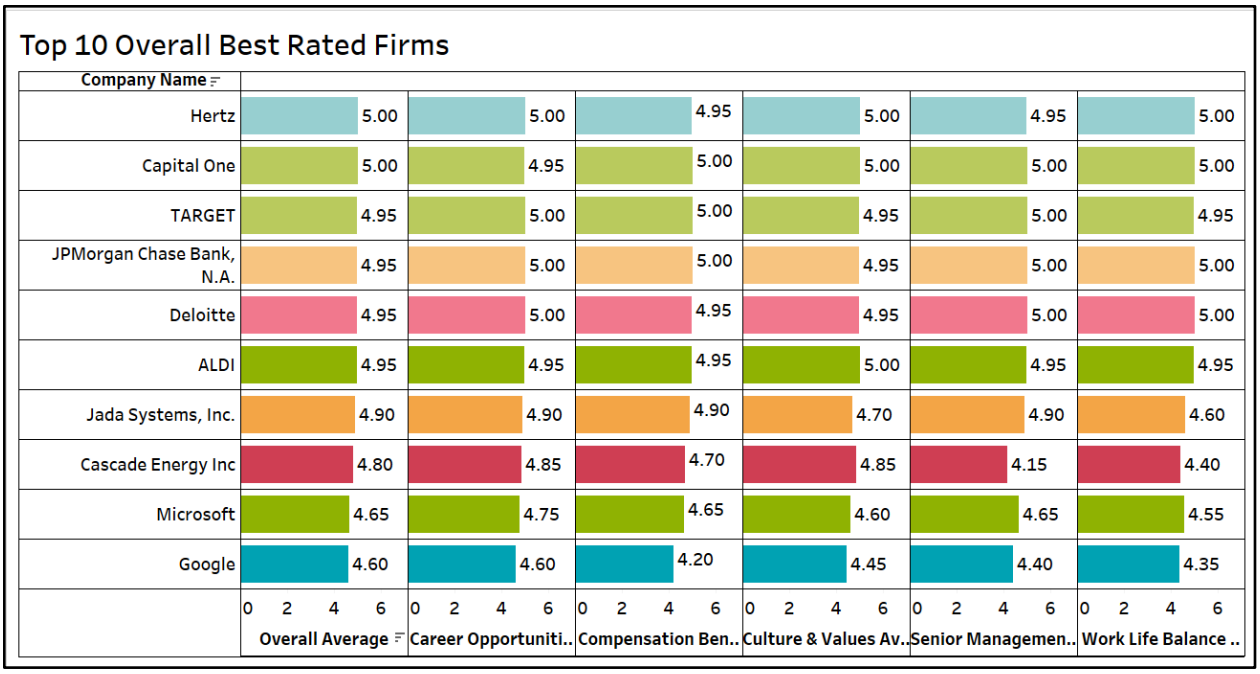


Figure 12: Tableau Output for Business Transaction 3

BUSINESS TRANSACTION 4: ***What is the percentage of remote vs onsite job opportunities?***

SQL CODE:

```
SELECT  ROUND(CAST(COUNT(CASE WHEN j.jobLocation = 'Remote' THEN 1 END) AS FLOAT) / CAST(COUNT(j.jobLocation) AS FLOAT) *100,2)
        AS 'Remote %',
        ROUND(CAST(COUNT(CASE WHEN j.jobLocation <> 'Remote' THEN 1 END) AS FLOAT) / CAST(COUNT(j.jobLocation) AS FLOAT) *100,2)
        AS 'Onsite %'
FROM [PerfectAnalyst.Job] j
```

Figure 13: SQL Code for Business Transaction 4

SQL OUTPUT:

	Remote %	Onsite %
1	45.98	54.02

Figure 14: SQL Output for Business Transaction 4

TABLEAU:

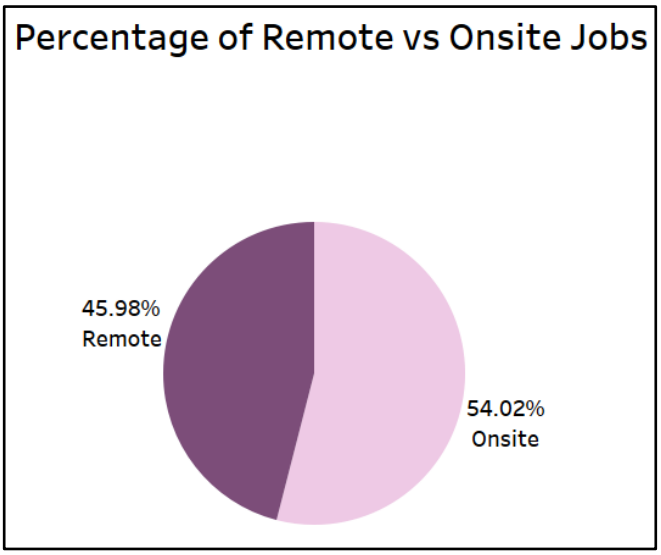


Figure 15: Tableau Output for Business Transaction 4

BUSINESS TRANSACTION 5: ***What are the companies with openings in multiple job functions?***

SQL CODE:

```
SELECT j.jobFunction AS 'Job Function', a.compName AS 'Company Name'
FROM (SELECT c.compId, c.compName
      FROM [PerfectAnalyst.Job] j, [PerfectAnalyst.Company] c
      WHERE j.compId = c.compId
      GROUP BY c.compId, c.compName
      HAVING COUNT(j.jobFunction) > 1 ) a, [PerfectAnalyst.Job] j
WHERE a.compId = j.compId
```

Figure 16: SQL Code for Business Transaction 5

SQL OUTPUT:

Results | Messages

	Job Function	Company Name
1	Data Engineer	Capital One
2	Data Engineer	MatchPointe Group
3	Data Engineer	etrailer.com
4	Data Engineer	DISH
5	Data Engineer	Aviation Structure Repairs
6	Data Engineer	PCS Global Tech
7	Data Engineer	Hewlett Packard Enterprise
8	Data Engineer	State Farm

rhsmith.umd.edu,97... | AD\aishsada (71) | BUDT703_Project_0506_10 | 00:00:00 | 30 rows

Figure 17: SQL Output for Business Transaction 5

TABLEAU:

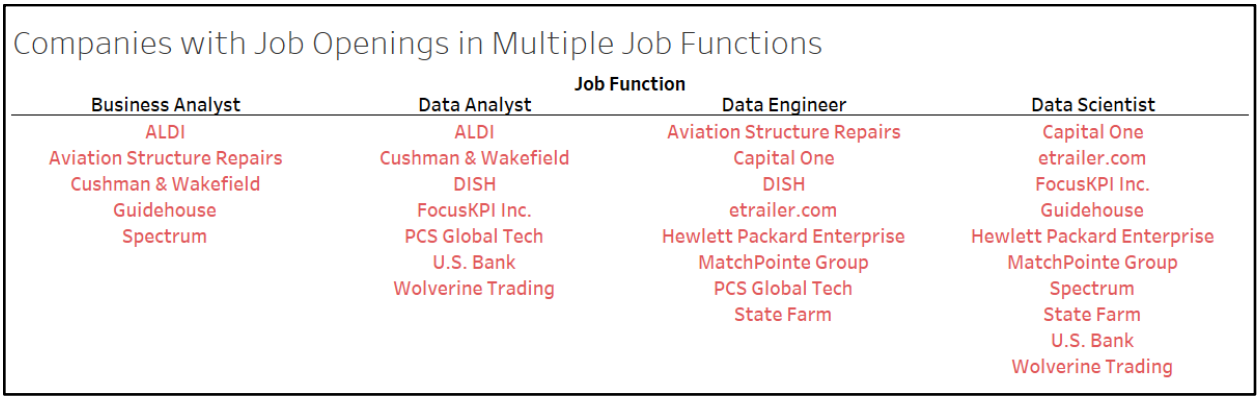


Figure 18: Tableau Output for Business Transaction 5

BUSINESS TRANSACTION 6: ***What is the profile match share by categories?***

SQL CODE:

```
SELECT  ROUND(CAST(COUNT(CASE WHEN p.pmCategory = 'Excellent' THEN 1 END) AS FLOAT) / CAST(COUNT(p.pmId) AS FLOAT) *100,2)
        AS 'Excellent %',
        ROUND(CAST(COUNT(CASE WHEN p.pmCategory = 'Average' THEN 1 END) AS FLOAT) / CAST(COUNT(p.pmId) AS FLOAT) *100,2)
        AS 'Average %',
        ROUND(CAST(COUNT(CASE WHEN p.pmCategory = 'Poor' THEN 1 END) AS FLOAT) / CAST(COUNT(p.pmId) AS FLOAT) *100,2)
        AS 'Poor %'
FROM [PerfectAnalyst.ProfileMatch] p
```

Figure 18: SQL Code for Business Transaction 6

SQL OUTPUT:

	Excellent %	Average %	Poor %
1	28.21	43.59	28.21

Figure 18: SQL Output for Business Transaction 6

TABLEAU:

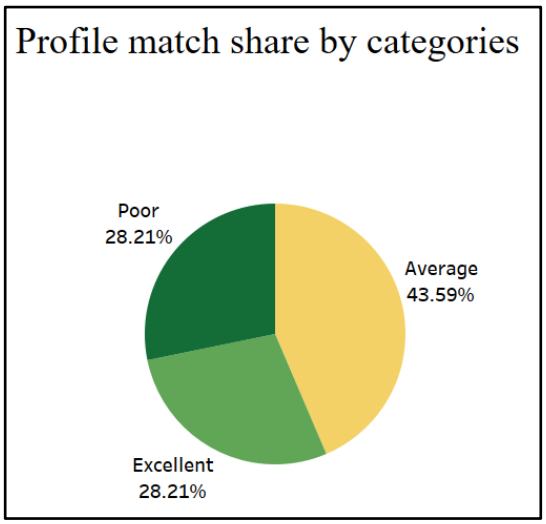
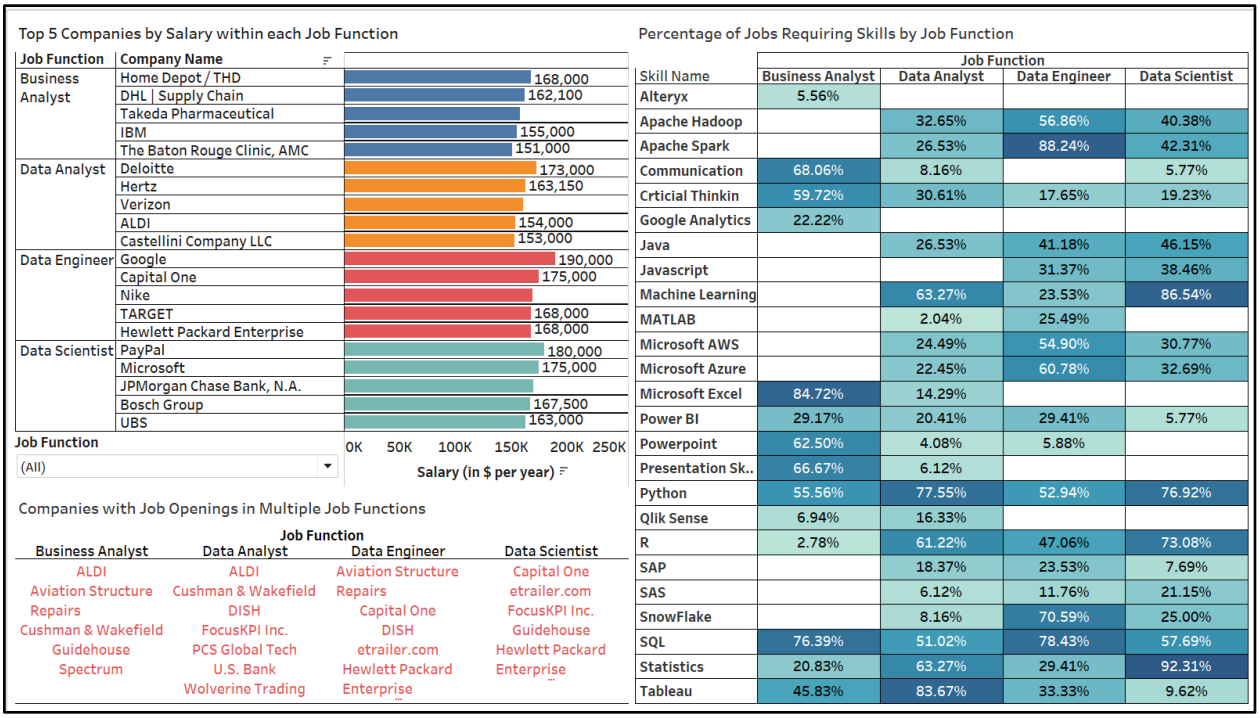


Figure 19: Tableau Output for Business Transaction 6

DASHBOARD:


Top 5 Companies by Salary within each Job Function
Job Function
Company Name
Business Analyst
Home Depot / THD 168,000
DHL | Supply Chain 162,100
Takeda Pharmaceutical
IBM 155,000
The Baton Rouge Clinic, AMC 151,000
Data Analyst
Deloitte 173,000
Hertz 163,150
Verizon
ALDI 154,000
Castellini Company LLC 153,000
Data Engineer
Google 190,000
Capital One 175,000
Nike
TARGET 168,000
Hewlett Packard Enterprise 168,000
Data Scientist
PayPal 180,000
Microsoft 175,000
JPMorgan Chase Bank, N.A.
Bosch Group 167,500
UBS 163,000
0K 50K 100K 150K 200K 250K
Salary (in $ per year)
Job Function
(All)
Companies with Job Openings in Multiple Job Functions
Job Function
Business Analyst
ALDI
Aviation Structure Repairs
Cushman & Wakefield
Guidehouse
Spectrum
Data Analyst
ALDI
Cushman & Wakefield
DISH
FocusKPI Inc.
PCS Global Tech
U.S. Bank
Wolverine Trading
Data Engineer
Aviation Structure Repairs
Capital One
DISH
etrailer.com
Hewlett Packard Enterprise
Data Scientist
Capital One
etrailer.com
FocusKPI Inc.
Guidehouse
Hewlett Packard Enterprise
Percentage of Jobs Requiring Skills by Job Function
Skill Name
Job Function
Business Analyst
Data Analyst
Data Engineer
Data Scientist
Alteryx 5.56%
Apache Hadoop 32.65% 56.86% 40.38%
Apache Spark 26.53% 88.24% 42.31%
Communication 68.06% 8.16% 5.77%
Crticial Thinkin 59.72% 30.61% 17.65% 19.23%
Google Analytics 22.22%
Java 26.53% 41.18% 46.15%
Javascript 31.37% 38.46%
Machine Learning 63.27% 23.53% 86.54%
MATLAB 2.04% 25.49%
Microsoft AWS 24.49% 54.90% 30.77%
Microsoft Azure 22.45% 60.78% 32.69%
Microsoft Excel 84.72% 14.29%
Power BI 29.17% 20.41% 29.41% 5.77%
Powerpoint 62.50% 4.08% 5.88%
Presentation Sk.. 66.67% 6.12%
Python 55.56% 77.55% 52.94% 76.92%
Qlik Sense 6.94% 16.33%
R 2.78% 61.22% 47.06% 73.08%
SAP 18.37% 23.53% 7.69%
SAS 6.12% 11.76% 21.15%
SnowFlake 8.16% 70.59% 25.00%
SQL 76.39% 51.02% 78.43% 57.69%
Statistics 20.83% 63.27% 29.41% 92.31%
Tableau 45.83% 83.67% 33.33% 9.62%