



# SQL Project with PostgreSQL

## PostgreSQL

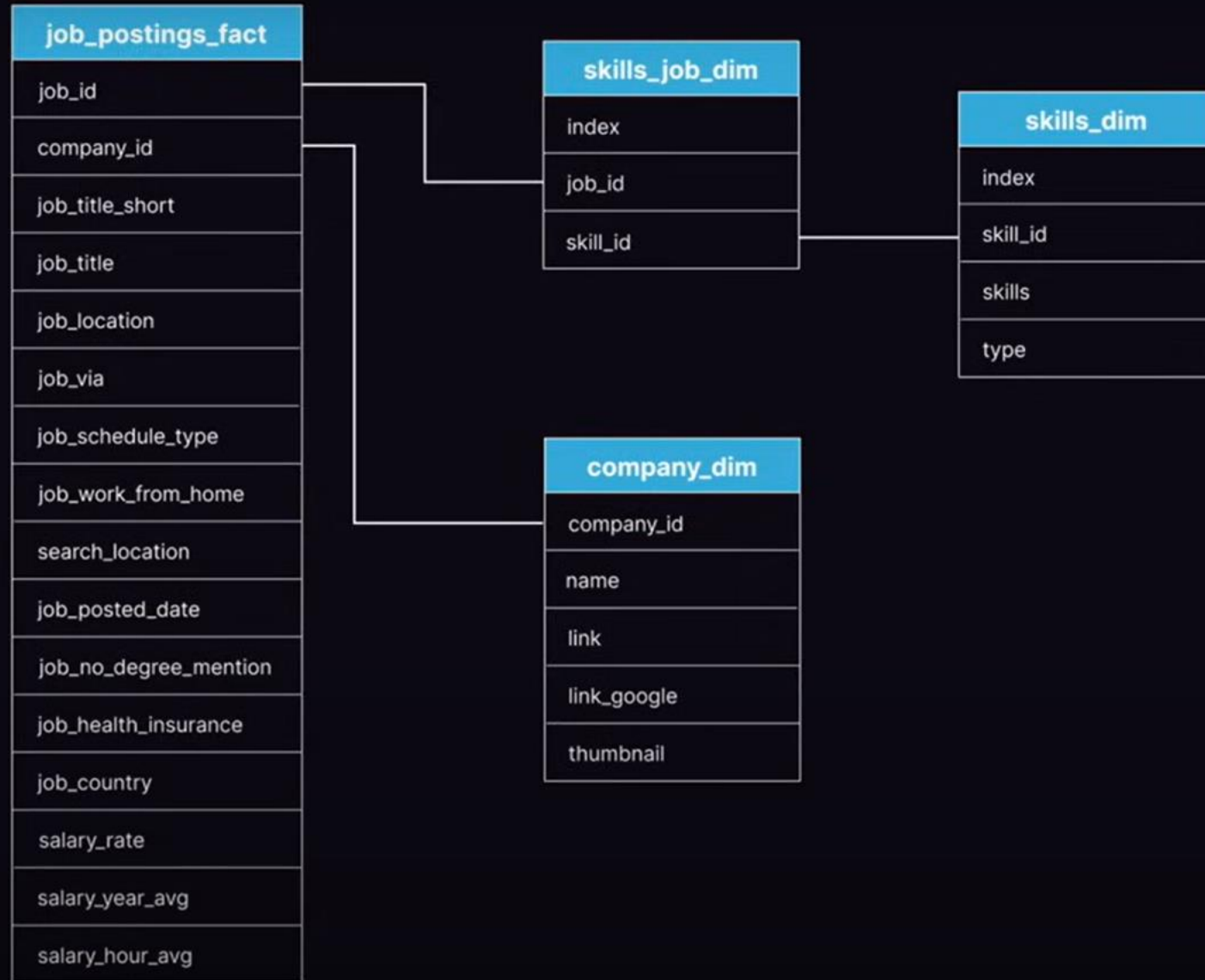
Finding the Suitable Job Over Various Job Platforms

Special Thanks :-

Luke Barousse

- Fardeen Ahmed Khan

## Database Structure – Depicting Relationship Between Tables and Keys Relating Them



### Question 1 - What are the Top Paying Data Analyst jobs?

- Identify the top 10 Highest paying Data Analyst Jobs available Remotely
- Focus on job postings with specifies Salary amount excluding nulls
- Why? Highlight the top-paying opportunities for Data Analyst

```
SELECT
    job_id,
    job_title,
    company_table.name as Company_Name,
    job_location,
    job_schedule_type,
    salary_year_avg,
    job_posted_date
from job_postings_fact as jobs_table
LEFT JOIN company_dim as company_table
ON jobs_table.company_id = company_table.company_id
where
job_title_short='Data Analyst' and job_location = 'Anywhere' and salary_year_avg IS NOT NULL
ORDER BY salary_year_avg DESC
LIMIT 10;
```

# The Top 10 Data Analyst Jobs Available for Work From Home i.e. Anywhere With Highest Salary In Desc Order.

<div> <div>1_top_paying_jobs.sql</div> <div> <div>sql_course: SELECT job...</div> <div>×</div> </div> </div>						
job_id	job_title	company_name	job_location	job_schedule_type	salary_year_avg	job_posted_date
abc Filter...	abc Filter...	abc Filter...	abc Filter...	abc Filter...	abc Filter...	abc Filter...
226942	Data Analyst	Mantys	Anywhere	Full-time	650000.0	2023-02-20 15:13:33
547382	Director of Analytics	Meta	Anywhere	Full-time	336500.0	2023-08-23 12:04:42
552322	Associate Director- Data Insights	AT&T	Anywhere	Full-time	255829.5	2023-06-18 16:03:12
99305	Data Analyst, Marketing	Pinterest Job Advertisements	Anywhere	Full-time	232423.0	2023-12-05 20:00:40
1021647	Data Analyst (Hybrid/Remote)	Uclahealthcareers	Anywhere	Full-time	217000.0	2023-01-17 00:17:23
168310	Principal Data Analyst (Remote)	SmartAsset	Anywhere	Full-time	205000.0	2023-08-09 11:00:01
731368	Director, Data Analyst - HYBRID	Inclusively	Anywhere	Full-time	189309.0	2023-12-07 15:00:13
310660	Principal Data Analyst, AV Performance...	Motional	Anywhere	Full-time	189000.0	2023-01-05 00:00:25
1749593	Principal Data Analyst	SmartAsset	Anywhere	Full-time	186000.0	2023-07-11 16:00:05
387860	ERM Data Analyst	Get It Recruit - Information Tech...	Anywhere	Full-time	184000.0	2023-06-09 08:01:04

**Question 2** - What skills are required for top paying Data Analyst jobs covered in previous Question

- Use the first query where we found top 10 data analyst remote jobs
- Add specific skills required for that role
- Why? It provide a detailed view skills for that high paying jobs helping job seekers in learning with skill can get them high paying jobs

```
with top_paying_jobs as (SELECT
    job_id,
    job_title,
    salary_year_avg,
    company_table.name as Company_Name
from job_postings_fact as jobs_table
LEFT JOIN company_dim as company_table
ON jobs_table.company_id =
company_table.company_id
where
    job_title_short='Data Analyst' and
job_location = 'Anywhere' and salary_year_avg IS
NOT NULL
ORDER BY salary_year_avg DESC
LIMIT 10)
```

```
SELECT
    top_paying_jobs.*,
    skills
from top_paying_jobs
INNER JOIN skills_job_dim as skills_to_jobs ON
top_paying_jobs.job_id = skills_to_jobs.job_id
INNER JOIN skills_dim as skills_name on
skills_to_jobs.skill_id = skills_name.skill_id
ORDER BY salary_year_avg DESC;
```

# Top 10 Data Analyst Jobs based on Highest Salary and their Required Skills as Job Specification

<div><div>2_top_paying_job_skills.sql</div><div>sql_course: with top_paying_... X</div></div>					<div><div>2_top_paying_job_skills.sql</div><div>sql_course: with top_paying_... X</div></div>				
job_id	job_title	salary_year_avg	company_name	skills	job_id	job_title	salary_year_avg	company_name	skills
<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>	<div>abc Filter...</div>
552322	Associate Director- Data Insights	255829.5	AT&T	sql	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	sql
552322	Associate Director- Data Insights	255829.5	AT&T	python	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	python
552322	Associate Director- Data Insights	255829.5	AT&T	r	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	go
552322	Associate Director- Data Insights	255829.5	AT&T	azure	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	snowflake
552322	Associate Director- Data Insights	255829.5	AT&T	databricks	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	pandas
552322	Associate Director- Data Insights	255829.5	AT&T	aws	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	numpy
552322	Associate Director- Data Insights	255829.5	AT&T	pandas	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	excel
552322	Associate Director- Data Insights	255829.5	AT&T	pyspark	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	tableau
552322	Associate Director- Data Insights	255829.5	AT&T	jupyter	168310	Principal Data Analyst (Remote)	205000.0	SmartAsset	gitlab
552322	Associate Director- Data Insights	255829.5	AT&T	excel	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	sql
552322	Associate Director- Data Insights	255829.5	AT&T	tableau	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	python
552322	Associate Director- Data Insights	255829.5	AT&T	power bi	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	azure
552322	Associate Director- Data Insights	255829.5	AT&T	powerpoint	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	aws
99305	Data Analyst, Marketing	232423.0	Pinterest Job Advertis...	sql	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	oracle
99305	Data Analyst, Marketing	232423.0	Pinterest Job Advertis...	python	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	snowflake
99305	Data Analyst, Marketing	232423.0	Pinterest Job Advertis...	r	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	tableau
99305	Data Analyst, Marketing	232423.0	Pinterest Job Advertis...	hadoop	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	power bi
99305	Data Analyst, Marketing	232423.0	Pinterest Job Advertis...	tableau	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	sap
1021647	Data Analyst (Hybrid/Remote)	217000.0	Uclahealthcareers	sql	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	jenkins
1021647	Data Analyst (Hybrid/Remote)	217000.0	Uclahealthcareers	crystal	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	bitbucket
1021647	Data Analyst (Hybrid/Remote)	217000.0	Uclahealthcareers	oracle	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	atlassian
1021647	Data Analyst (Hybrid/Remote)	217000.0	Uclahealthcareers	tableau	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	jira
1021647	Data Analyst (Hybrid/Remote)	217000.0	Uclahealthcareers	flow	731368	Director, Data Analyst - HYBRID	189309.0	Inclusively	confluence

### Question 3 - What are the most in demand skills for ad Data Analyst Role

- Identify the Top 10 skills for Data Analyst Job Position across all Platform and Companies
- Focus on All job postings
- Why? Retrieves the top 5 skills required to be a Data Analyst to be relevant in the job market

```
SELECT
    skills_to_jobs.skill_id,
    skills_table.skills,
    count(jobs_table.job_id) as Total_jobs
from job_postings_fact as jobs_table
INNER JOIN skills_job_dim as skills_to_jobs
ON jobs_table.job_id = skills_to_jobs.job_id
INNER JOIN skills_dim as skills_table
ON skills_to_jobs.skill_id = skills_table.skill_id
where job_title_short='Data Analyst'
GROUP BY skills_to_jobs.skill_id, skills
ORDER BY Total_jobs DESC
LIMIT 10;
```

## Top 10 Skills Required to be a Data Analyst which are high in Demand With Job Counts Across all Platform

3_top_demanded_skills.sql ● sql_course: SELECT ... ✕		
skill_id	skills	total_jobs
abc Filter...	abc Filter...	abc Filter...
0	sql	92628
181	excel	67031
1	python	57326
182	tableau	46554
183	power bi	39468
5	r	30075
7	sas	14034
186	sas	14034
196	powerpoint	13848
188	word	13591



**Question 4 :** What are the top paying skills based on salary

- Look at the average salary associated with each skill for Data Analyst Position
- Focuses on roles based on Salaries regardless of location
- Why? It defines how different skills impact the Overall Salary for a job person and to find most financially rewarding skill to learn and improve

```
SELECT
    skills_table.skill_id,
    skills_table.skills,
    ROUND(AVG(salary_year_avg),0) as avg_salary
from job_postings_fact AS jobs_table
INNER JOIN skills_job_dim as skills_to_job
ON jobs_table.job_id = skills_to_job.job_id
INNER JOIN skills_dim as skills_table
ON skills_to_job.skill_id = skills_table.skill_id
WHERE job_title_short = 'Data Analyst' AND salary_year_avg IS NOT NULL
GROUP BY skills_table.skill_id, skills
ORDER BY avg_salary DESC
LIMIT 20;
```

These Are the Top Skills to be a Data Analyst not on the basis of Job Count but what is the Average Salary One Might expect or what value this particular skill add to some one Package.  
It is important to assess the financial metric of all skills important in Data Analyst Domain.

4_top_paying_skills.sql	sql_course: SELECT skill...	
skill_id	skills	avg_salary
Filter...	Filter...	Filter...
224	svn	400000
38	solidity	179000
65	couchbase	160515
206	datarobot	155486
27	golang	155000
109	mxnet	149000
119	dplyr	147633
73	vmware	147500
212	terraform	146734
250	twilio	138500
220	gitlab	134126
98	kafka	129999
222	puppet	129820
100	keras	127013
101	pytorch	125226
31	perl	124686
223	ansible	124370
121	hugging face	123950
99	tensorflow	120647
63	cassandra	118407

**Question 5 :** What are the best skills to learn aka best in demand and high in paying for Remote Jobs

- Identify skills in high demand and with high average salaries for Data Analyst Roles
- Why? Target skills offer job security as they are high in demand and financially They are sound as well as being high in salaries for career development in data analysis

```
WITH skills_demand AS (SELECT
skills_table.skill_id,
    skills_table.skills,
    count(jobs_table.job_id) as total_jobs
from job_postings_fact as jobs_table
INNER JOIN skills_job_dim as skills_to_jobs
ON jobs_table.job_id = skills_to_jobs.job_id
INNER JOIN skills_dim as skills_table
ON skills_to_jobs.skill_id =
skills_table.skill_id
WHERE job_title_short = 'Data Analyst'
AND salary_year_avg IS NOT NULL
AND job_work_from_home = TRUE
GROUP BY skills_table.skill_id, skills
), ----→
```

```
average_salary as(
SELECT skills_table.skill_id,
    skills_table.skills,
    ROUND(AVG(salary_year_avg),0) as
avg_salary
from job_postings_fact AS jobs_table
INNER JOIN skills_job_dim AS skills_to_jobs
ON jobs_table.job_id = skills_to_jobs.job_id
INNER JOIN skills_dim as skills_table
ON skills_to_jobs.skill_id =
skills_table.skill_id
WHERE job_title_short = 'Data Analyst'
AND salary_year_avg IS NOT NULL
AND job_work_from_home = TRUE
GROUP BY skills_table.skill_id, skills)
```

## Continued from Previous Slide

```
WITH skills_demand AS (SELECT
    skills_table.skill_id,
    skills_table.skills,
    count(jobs_table.job_id) as total_jobs
from job_postings_fact as jobs_table
INNER JOIN skills_job_dim as skills_to_jobs
ON jobs_table.job_id = skills_to_jobs.job_id
INNER JOIN skills_dim as skills_table
ON skills_to_jobs.skill_id =
skills_table.skill_id
WHERE job_title_short = 'Data Analyst'
AND salary_year_avg IS NOT NULL
AND job_work_from_home = TRUE
GROUP BY skills_table.skill_id, skills
), ---->
```

```
average_salary as(
    SELECT skills_table.skill_id,
    skills_table.skills,
    ROUND(AVG(salary_year_avg),0) as avg_salary
from job_postings_fact AS jobs_table
INNER JOIN skills_job_dim AS skills_to_jobs
ON jobs_table.job_id = skills_to_jobs.job_id
INNER JOIN skills_dim as skills_table
ON skills_to_jobs.skill_id =
skills_table.skill_id
WHERE job_title_short = 'Data Analyst'
AND salary_year_avg IS NOT NULL
AND job_work_from_home = TRUE
GROUP BY skills_table.skill_id, skills)
```

```
SELECT skills_demand.*,
    average_salary.avg_salary
from skills_demand
INNER JOIN average_salary
on skills_demand.skill_id =
average_salary.skill_id
ORDER BY total_jobs DESC;
```

The Idea here is to find the skills with most Job count and then Find the avg salary of the Relevant skill so that It can be easy for Aspiring Data Analyst to Choose the Learning Point which can give them a chance to start working in the Data Domain and at the same Time can provide enough Financial Stability as well.

5_optimal_skills.sql • sql_course: WITH skills_dema... X			
skill_id	skills	total_jobs	avg_salary
0	sql	398	97237
181	excel	256	87288
1	python	236	101397
182	tableau	230	99288
5	r	148	100499
183	power bi	110	97431
7	sas	63	98902
186	sas	63	98902
196	powerpoint	58	88701
185	looker	49	103795
188	word	48	82576
79	oracle	37	104534
80	snowflake	37	112948
61	sql server	35	97786
74	azure	34	111225
192	sheets	32	86088
76	aws	32	108317
215	flow	28	97200
8	go	27	115320
199	spss	24	92170
22	vba	24	88783
97	hadoop	22	113193
233	jira	20	104918

WORK FROM HOME IS TRUE

5_optimal_skills.sql • sql_course: WITH skills_dema... X			
skill_id	skills	total_jobs	avg_salary
233	jira	20	104918
9	javascript	20	97587
195	sharepoint	18	81634
201	alteryx	17	94145
4	java	17	106906
78	redshift	16	99936
197	ssrs	14	99171
187	qlik	13	99631
198	outlook	13	90077
2	nosql	13	101414
77	bigquery	13	109654
92	spark	13	99077
194	ssis	12	106683
13	c++	11	98958
234	confluence	11	114210
75	databricks	10	141907
184	dax	10	104500
14	c#	10	86540
204	visio	10	95842
56	mysql	9	95224
16	t-sql	9	96365
26	c	9	98938
93	pandas	9	151821

# Conclusion

It was Done on Real Job Posting Collected from Various Websites and Platform So it is a Real Dataset.

The Data Analyst Job Market is Expanding and It is the right time to Upskill Yourself

For More Info Watch This Video

<https://www.youtube.com/watch?v=7mz73uXD9DA>