SQL Project: Study Case Unicorn Companies

By: Iing Fitria Kurniasari

Background

A **unicorn** is a privately held startup company with a current valuation of US$1 billion or more. This dataset contains companies classified as unicorns in early April 2022, so startup that are no longer classified unicorns for reasons of IPO or mergers are not included in this dataset.

Overview Dataset

unicorn_companies **table**

`SELECT * FROM unicorn_companies`

	company_id integer	company text	city text	country text	continent text
1	189	Otto Bock…	Duderstadt	Germany	Europe
2	848	Matrixport	[null]	Singapore	Asia
3	556	Cloudinary	Santa Clara	United St…	North Am…
4	999	PLACE	Bellingham	United St…	North Am…
5	396	candy.com	New York	United St…	North Am…

Total rows: 1000 of 1074   Query complete 00:00:03.424

This data contains of 1074 rows.

unicorn_industries **table**

`SELECT * FROM unicorn_industries`

	company_id integer	industry text
1	189	Health
2	848	Fintech
3	556	Internet s…
4	999	Internet s…
5	396	Fintech

Total rows: 1000 of 1074   Query complete 00:00:00.425

This table contains the type of industries for each company in unicorn_industries table.

Overview Dataset

unicorn_funding **table**

SELECT * FROM unicorn_funding

	company_id integer 🔒	valuation bigint 🔒	funding bigint 🔒	select_investors text 🔒
1	189	4000000000	0	EQT Partners
2	848	1000000000	100000000	"Dragonfly Capti…
3	556	2000000000	100000000	"Blackstone, Be…
4	999	1000000000	100000000	"Goldman Sachs…
5	396	2000000000	100000000	"Insight Partner…

Total rows: 1000 of 1074   Query complete 00:00:00.292

unicorn_dates **table**

SELECT * FROM unicorn_dates

	company_id integer 🔒	date_joined date 🔒	year_founded integer 🔒
1	189	2017-06-24	1919
2	848	2021-06-01	2019
3	556	2022-02-15	2011
4	999	2021-11-17	2020
5	396	2021-10-21	2021

Total rows: 1000 of 1074   Query complete 00:00:01.567

1 Using unicorn_companies data, perform a query to get the number of unicorn companies for each continent, ordered by total company descendingly.

Query

```
SELECT
    continent,
    COUNT(company_id) as total_company
FROM unicorn_companies
GROUP BY 1
ORDER BY 2 DESC
```

Result

	continent 🔒 text	total_company 🔒 bigint
1	North America	589
2	Asia	310
3	Europe	143
4	South America	21
5	Oceania	8
6	Africa	3

North America has the most number of unicorn companies.

2 Using unicorn_companies data, perform a query to display the country that has more than 100 unicorns.

Query

```sql
SELECT
    country,
    COUNT(company_id) as total_unicorn
FROM unicorn_companies
GROUP BY 1
HAVING COUNT(company_id) > 100;
```

Result

	country text	total_unicorn bigint
1	China	173
2	United States	562

Only **China and United States** has more than 100 unicorns.

3 Using unicorn_companies and unicorn_funding data, perform a query to get records of the largest industry based on total funding and its average valuation.

Query

Result

```sql
SELECT
    industry,
    SUM(funding) as total_funding,
    ROUND(AVG(valuation),0) as average_of_valuation
FROM unicorn_industries i
JOIN unicorn_funding f
ON i.company_id = f.company_id
GROUP BY 1
ORDER BY 2 DESC
LIMIT 1;
```

	industry text	total_funding numeric	average_of_valuation numeric
1	Fintech	107996000000	3937500000

Fintech is the largest industry that has total funding 107996 million and the average valuation reaches 3937.5 million.

4 Using unicorn_companies, unicorn_industries, and unicorn_dates data, perform a query to get the total company for the Fintech industry grouped by year.

Query

```sql
SELECT
    EXTRACT(YEAR FROM d.date_joined) as year,
    COUNT(d.company_id) as total_company
FROM unicorn_industries i
JOIN unicorn_dates d
ON i.company_id = d.company_id
WHERE industry = 'Fintech' AND EXTRACT(YEAR FROM d.date_joined) BETWEEN 2016 AND 2022
GROUP BY 1;
```

Result

	year numeric	total_company bigint
1	2017	6
2	2018	10
3	2019	20
4	2020	15
5	2021	138
6	2022	31

5.a Using unicorn_companies, unicorn_industries, and unicorn_funding data, perform a query to display company detail data from the country whose largest valuation.

Query

```sql
SELECT
    company, city, country, continent, industry, valuation
FROM unicorn_companies s
JOIN unicorn_industries i
ON s.company_id = i.company_id
JOIN unicorn_funding f
ON s.company_id = f.company_id
WHERE valuation = (SELECT max(valuation) FROM unicorn_funding)
```

Result

	company text	city text	country text	continent text	industry text	valuation bigint
1	Bytedance	Beijing	China	Asia	Artificial i...	180000000000

5.b Using unicorn_companies, unicorn_industries, and unicorn_funding data, perform a query to display company detail data from Indonesia that have the largest valuation.

Query

```sql
SELECT
    company, city, country, continent, industry, valuation
FROM unicorn_companies s
JOIN unicorn_industries i
ON s.company_id = i.company_id
JOIN unicorn_funding f
ON s.company_id = f.company_id
WHERE country = 'Indonesia'
ORDER BY valuation DESC
LIMIT 1
```

Result

	company text	city text	country text	continent text	industry text	valuation bigint
1	J&T Express	Jakarta	Indonesia	Asia	"Supply chain, logi...	20000000000

6 Using unicorn_companies and unicorn_dates data, perform a query to display the age of the old company when it merged to become a unicorn company. Which country the company come from?

Query

```sql
SELECT
    company,
    country,
    (EXTRACT(YEAR FROM date_joined) - year_founded) as age_at_joining
FROM unicorn_companies s
JOIN unicorn_dates d
ON s.company_id = d.company_id
ORDER BY 3 DESC
LIMIT 1
```

Result

	company text	country text	age_at_joining numeric
1	Otto Bock HealthCare	Germany	98

7 Using unicorn_companies and unicorn_dates data, perform a query to display the age of the old company when it merged to become a unicorn company and from which country the company originated. Only for a company that founded between 1960 and 2000.

Query

```
WITH age AS (
SELECT
    company,
    country,
    (EXTRACT(YEAR FROM date_joined) - year_founded) as age_at_joining,
    EXTRACT(MONTH FROM date_joined) as month_joined
FROM unicorn_companies s
    JOIN unicorn_dates d
    ON s.company_id = d.company_id
WHERE year_founded BETWEEN 1960 AND 2000
ORDER BY 3 DESC
LIMIT 1)
SELECT company, country, age_at_joining
FROM age
ORDER BY age_at_joining DESC, month_joined ASC
```

Result

	company text	country text	age_at_joining numeric
1	Five Star Business Finance	India	37

Total rows: 1 of 1   Query complete 00:00:00.161

8.a Using unicorn_companies, unicorn_industries, and unicorn_funding data, perform a query to get the number of companies are financed by at least one investor with the name 'venture'.

Query

```sql
SELECT
    COUNT(company_id)
FROM unicorn_funding
WHERE LOWER(select_investors) like '%venture%'
```

Result

	count bigint 🔒
1	603

Total rows: 1 of 1

8.b Using unicorn_companies, unicorn_industries, and unicorn_funding data, perform a query to get the number of companies are financed by at least on investor with the names contain the letter 'venture', 'capital', or 'partner'.

Query

```sql
SELECT
    COUNT(DISTINCT CASE WHEN LOWER(select_investors) LIKE '%venture%' THEN company_id END) AS investor_venture,
    COUNT(DISTINCT CASE WHEN LOWER(select_investors) LIKE '%capital%' THEN company_id END) AS investor_capital,
    COUNT(DISTINCT CASE WHEN LOWER(select_investors) LIKE '%partner%' THEN company_id END) AS investor_partner
FROM unicorn_funding
```

Result

	investor_venture 🔒 bigint	investor_capital 🔒 bigint	investor_partner 🔒 bigint
1	603	611	398

Total rows: 1 of 1   Query complete 00:00:00.129

9 Using unicorn_companies and unicorn_industries data, perform a query to display the number of logistics companies are unicorns in Asia and Indonesia.

Query

```sql
SELECT
    COUNT(DISTINCT(i.company_id)) AS total_Asia_logistics,
    COUNT(DISTINCT CASE WHEN country like '%Indonesia%' THEN country END) AS total_Indonesia_logistics
FROM unicorn_companies s
    JOIN unicorn_industries i
    ON s.company_id = i.company_id
WHERE industry like '%logistic%'
and continent like '%Asia%'
```

Result

	total_asia_logistics 🔒 bigint	total_indonesia_logistics 🔒 bigint
1	26	1

Total rows: 1 of 1   Query complete 00:00:00.951

10 In Asia, there are three countries with the highest number of unicorns. Using unicorn_companies and unicorn_industries data, perform a query to show the number of unicorns in each industry and country of origin in Asia, except of these three countries. Sort by industry, number of companies (decreasing), and country of origin.

Query

```sql
WITH top3_asia AS (
SELECT
    country,
    COUNT(s.company_id) as total_unicorn
FROM unicorn_companies s
WHERE continent = 'Asia'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 3)
SELECT
    industry,
    country,
    COUNT(DISTINCT s.company_id) as total_unicorn
FROM unicorn_companies s
    JOIN unicorn_industries i
    ON s.company_id = i.company_id
WHERE country NOT IN (SELECT country FROM top3_asia)
AND continent = 'Asia'
GROUP BY 1,2
ORDER BY 3 DESC, 1,2;
```

Result

	industry text	country text	total_unicorn bigint
1	E-commerce & direct-to-consum…	Singapore	3
2	E-commerce & direct-to-consum…	South Korea	3
3	Fintech	Singapore	3
4	Artificial intelligence	Singapore	2
5	Fintech	Hong Kong	2

Total rows: 39 of 39   Query complete 00:00:04.056

11 The United States, China, and India are the three countries with the most unicorns. Using unicorn_companies and unicorn_industries data, perform a query to get the type of industry that doesn't originate in India.

Query

```sql
SELECT
    DISTINCT industry
FROM unicorn_industries i
WHERE i.industry NOT IN (
    SELECT
        DISTINCT industry
    FROM unicorn_companies s
        JOIN unicorn_industries i1
        ON s.company_id = i1.company_id
    WHERE s.country = 'India'
)
```

Result

	industry text 🔒	
1	Consumer & retail	
2	Hardware	
3	Artificial intelligence	
4	Cybersecurity	

Total rows: 4 of 4   Query complete 00:00:00.249

12 Using unicorn_companies, unicorn_dates, and unicorn_funding data, perform a query to find the three industries with the most unicorns in 2019-2021 and display the number of unicorns and their average valuation (in billions) each year.

Query

```sql
WITH top_3 AS (
SELECT
    i.industry,
    COUNT(DISTINCT i.company_id)
FROM unicorn_industries i
    INNER JOIN unicorn_dates d
    ON i.company_id = d.company_id
WHERE
    EXTRACT(YEAR FROM d.date_joined) IN (2019,2020,2021)
GROUP BY 1
ORDER BY 2 DESC
LIMIT 3
)
SELECT
    i.industry,
    EXTRACT(YEAR FROM d.date_joined) AS year_joined,
    COUNT(DISTINCT i.company_id) AS total_company,
    ROUND(AVG(f.valuation)/1000000000,2) AS avg_valuation_billion
FROM unicorn_industries i
INNER JOIN unicorn_funding f
    ON i.company_id = f.company_id
INNER JOIN (SELECT * FROM unicorn_dates WHERE EXTRACT(YEAR FROM date_joined) IN (2019,2020,2021)) d
    ON i.company_id = d.company_id
WHERE i.industry IN (SELECT industry FROM top_3)
GROUP BY 1,2
```

12 Using unicorn_companies, unicorn_dates, and unicorn_funding data, perform a query to find the three industries with the most unicorns in 2019-2021 and display the number of unicorns and their average valuation (in billions) each year.

Result

	industry text	year_joined numeric	total_company bigint	avg_valuation_billion numeric
1	E-commerce & direct-to-consu…	2021	47	2.47
2	E-commerce & direct-to-consu…	2020	16	4.00
3	E-commerce & direct-to-consu…	2019	12	2.58
4	Fintech	2021	138	2.75
5	Fintech	2020	15	4.33
6	Fintech	2019	20	6.80
7	Internet software & services	2021	119	2.15
8	Internet software & services	2020	20	4.35
9	Internet software & services	2019	13	4.23

Total rows: 9 of 9   Query complete 00:00:00.249

13 Using unicorn_companies, unicorn_dates, and unicorn_funding data, perform a query to get the number of unicorn companies for each continent, ordered by total company descendingly.

Query

```sql
WITH country_level AS (
SELECT
    s.country,
    COUNT(DISTINCT s.company_id) AS total_per_country
FROM unicorn_companies s
GROUP BY 1
)
SELECT
    *,
    CONCAT(ROUND((total_per_country / SUM(total_per_country) OVER())*100,2),'%') AS pct_company
FROM country_level
ORDER BY 2 DESC;
```

13 Using unicorn_companies, unicorn_dates, and unicorn_funding data, perform a query to get the number of unicorn companies for each continent, ordered by total company descendingly.

Result

	country text	total_per_country bigint	pct_company text
1	United States	562	52.33%
2	China	173	16.11%
3	India	65	6.05%
4	United Kingdom	43	4.00%
5	Germany	26	2.42%
6	France	24	2.23%
7	Israel	20	1.86%
8	Canada	19	1.77%
9	Brazil	16	1.49%
10	South Korea	12	1.12%

Contact me at :

in ling Fitria Kurniasari

M fitriakurniasari8@gmail.com

For the detail processed check on my medium :
https://medium.com/@fitriakurniasari8/data-analysis-with-sql-6a10b683617