DATA ANALYSIS

We are given this dataset for the Tech Company Fundings in 2020 and we are given several question to analyses that data using SQL:

1. What Company got the most funding?
2. What's the most popular bussiness sector to invest?
3. Which Country got the most funding? And in what sector is the biggest funding there?
4. What funding stage are the most in the table?
5. Can you prove how the stage funding can affect the average of funding the company get?

You can check the full datasets in this link below :

https://www.kaggle.com/datasets/shivamb/tech-company-fundings-2020-onwards

	ID	COMPANY	WEBSITE	REGION	VERTICAL	FUNDING_AMOUNT_USD	FUNDING_STAGE	FUNDING_DATE
1	83	Funnel	https://funnel.io/	Sweden	B2B Software	47000000	Series B	20-Jan
2	84	Justworks	http://www.justworks.com	Unit...	B2B Software	50000000	Series E	20-Jan
3	85	SECURITI.ai	https://www.securiti.ai/	Unit...	Artificial Intelligence	50000000	Series B	20-Jan
4	86	LaunchDarkly	http://launchdarkly.com	Unit...	B2B Software	54000000	Series D	20-Jan
5	87	SiteMinder	https://www.siteminder.com/	Aust...	B2B Software	70000000	Series C	20-Jan
6	88	Personio	https://www.personio.com	Germany	Human Resources	75000000	Series C	20-Jan
7	89	Element Biosciences	https://www.elementbioscien...	Unit...	Biotechnology	80000000	Series B	20-Jan
8	90	Verkada	https://www.verkada.com/	Unit...	Computer Vision	80000000	Series C	20-Jan
9	91	Bounce	https://bounceshare.com/	India	Transportation	97500000	Series D	20-Jan
10	92	ActiveCampaign	http://www.activecampaign.com	Unit...	B2B Software	100000000	Series B	20-Jan
11	93	Livspace	http://www.livspace.com/	India	Marketplace	100000000	Series D	20-Jan
12	94	Sisense	http://www.sisense.com	Unit...	B2B Software	100000000	Series F	20-Jan
13	95	Qonto	https://www.qonto.eu	France	FinTech	115000000	Series C	20-Jan
14	96	Joby Aviation	http://www.jobyaviation.com/	Unit...	Aerospace	590000000	Series C	20-Jan
15	97	Gojek	https://www.gojek.io/	Indo...	Food Delivery	1200000000	Series G	20-Mar
16	98	Impossible Foods	https://impossiblefoods.com/	Unit...	Food and Beverage	500000000	Series F	20-Mar
17	99	Bakkt	https://www.bakkt.com	Unit...	Blockchain	300000000	Series B	20-Mar
18	100	Lilium	https://lilium.com/	Germany	Air Transportation	240000000	Series C	20-Mar
19	101	Colonies	https://www.livecolonies.com/	France	Co-Living	200000000	Series B	20-Mar
20	102	Via	https://www.ridewithvia.com	Unit...	Automotive	200000000	Series E	20-Mar
21	103	HashiCorp	https://www.hashicorp.com	Unit...	Cyber Security	175000000	Series E	20-Mar
22	104	ElevateBio	https://www.elevate.bio/	Unit...	Biotechnology	170000000	Series B	20-Mar
23	105	InSightec	https://www.insightec.com	Israel	Health Care	150000000	Series F	20-Mar
24	106	UserTesting	https://www.usertesting.com	Unit...	B2B Software	100000000	Series C	20-Mar
25	107	Lyra Health	https://www.lyrahealth.com	Unit...	Employee Benefits	75000000	Series C	20-Mar
26	108	Rigetti Computing	https://www.rigetti.com/	Unit...	Quantum Computing	71000000	Series D	20-Mar
27	109	Everlaw	https://everlaw.com/	Unit...	Legal	62000000	Series C	20-Mar
28	110	Axonius	https://www.axonius.com/	Israel	Cyber Security	58000000	Series C	20-Mar
29	111	Netlify	https://www.netlify.com	Unit...	Web Development	53000000	Series C	20-Mar
30	112	SpotOn	https://www.spoton.com	Unit...	FinTech	50000000	Series B	20-Mar

So, without further ado let's start our analyses.

1. **What company got the most funding?**

   To Answer the question we must know first what type of data using in that table.

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	ID	NUMBER(38,0)	No	(null)	1	(null)
2	COMPANY	VARCHAR2(29 BYTE)	No	(null)	2	(null)
3	WEBSITE	VARCHAR2(37 BYTE)	No	(null)	3	(null)
4	REGION	VARCHAR2(26 BYTE)	Yes	(null)	4	(null)
5	VERTICAL	VARCHAR2(27 BYTE)	No	(null)	5	(null)
6	FUNDING_AMOUNT_USD	VARCHAR2(11 BYTE)	No	(null)	6	(null)
7	FUNDING_STAGE	VARCHAR2(21 BYTE)	No	(null)	7	(null)
8	FUNDING_DATE	VARCHAR2(6 BYTE)	No	(null)	8	(null)

Turns out, most of the columns using the varchar type of data. If we want to use the MAX function to answer the question it will cause error because the MAX function can only use for the numeric type of data for example Float or Integer. If we want to change the type of data we have to empty the value in that column first. So one of the way to answer this question, we use this query below:

```
select COMPANY, FUNDING_AMOUNT_USD
from MYTABLE
where LENGTH(FUNDING_AMOUNT_USD)=(
select MAX(LENGTH(FUNDING_AMOUNT_USD))
from MYTABLE);
```

Here, we add the Length function after the MAX function in order to find the most character in that column FUNDING_AMOUNT_USD. The result is WestConnex are the company who got the most funding in the datasets we got.

	COMPANY	FUNDING_AMOUNT_USD
1	WestConnex	16600000000

2. What's the most popular business sector to invest?

To answer this question we can use the Count function in order to know how much the unique value that are in vertical column. After that we group by the vertical column so the data will be aggregated based on unique values in the vertical column. To have a better read for the answer we can order the result using the descending order based on the count.

```
select VERTICAL, count(*) AS VERTICAL_COUNT
from MYTABLE
group by VERTICAL
having COUNT(*)>1
order by VERTICAL_COUNT DESC;
```

The result shows that B2B Software are the most Bussiness Sector in the datasets. In this case, the difference in numbers between the B2B Software and the Cloud Computing sector is also relatively large by 300 more differences.

	VERTICAL	VERTICAL_COUNT
1	B2B Software	632
2	Cloud Computing	288
3	Artificial Intelligence	283
4	Finance	254
5	Blockchain	244
6	Cyber Security	147
7	Health Care	139
8	FinTech	104
9	Education	104
10	E-Commerce Store	93
11	Food and Beverage	89
12	Gaming	66
13	Marketplace	65
14	Human Resources	64
15	Real Estate	59

3. **Which country are the most in the datasets? And in what sector is the popular funding there?**

For this question we can answer by step by step. The first step is we want to know which country are the most found in the datasets. Similar with the step number 2, we can use the Count Function to counts the occurrences in the column Region. We use 'group by' so the data will be aggregated based on unique values in the region column and use 'Having' to applies a condition to filter the grouped data. We can use 'order by' to order the result in descending order.

```
select REGION, count(*) AS COUNTRY_COUNT
from MYTABLE
group by REGION
having COUNT(*)>1
order by COUNTRY_COUNT DESC;
```

The result show that the most country in the datasets is United States with 2033 Company, followed by United kingdom with 317 Company, and India with 155 Company.

	REGION	COUNTRY_COUNT
1	United States	2033
2	United Kingdom	317
3	India	155
4	Canada	111
5	Germany	101
6	France	95
7	Israel	93
8	Australia	53
9	Singapore	50
10	Spain	42
11	The Netherlands	35
12	Switzerland	30
13	Sweden	28
14	Finland	27
15	Nigeria	24

After we know that the United States are the most country we found on the datasets, we move on to know what business sector is the most popular to get the funding there. Similar query from before we can use the count function and using the 'group by' and 'order by' query. For additional we can filter using the where statement in column Region so that it can only select the data where the region is in United States.

```
SELECT VERTICAL, COUNT(*) AS SECTOR_COUNT
FROM MYTABLE
WHERE REGION = 'United States'
GROUP BY VERTICAL
ORDER BY SECTOR_COUNT DESC;
```

The result show that the B2B software are the most popular bussiness sector in United states, followed by Cloud Computing in the 2nd and Artificial Intelligence in the 3rd place. From this result we can also conclude that the three most sector bussiness in United States have similarities in the realm of IT.

	VERTICAL	SECTOR_COUNT
1	B2B Software	404
2	Cloud Computing	181
3	Artificial Intelligence	168
4	Blockchain	118
5	Finance	109
6	Health Care	100
7	Cyber Security	96
8	FinTech	55
9	Food and Beverage	54
10	Education	45
11	Real Estate	39
12	E-Commerce Store	38
13	Recruiting	37
14	Gaming	36
15	Human Resources	35

4. **What funding stage are the most in the table?**

   Pretty similar tasks with the previous number above, the difference is on the select statement. In this case we are using the funding_stage column to find the result.

```
select FUNDING_STAGE, count(*) as FUNDING_STAGE_COUNT
from MYTABLE
group by FUNDING_STAGE
having count(*)>1
order by FUNDING_STAGE_COUNT desc;
```

   The result show that funding stage Series A is the most numerous from the table with 951 company on this stage per 2020. Followed by Seed with 850 company and Series B with 648 company. From the result we can also conclude that majority of the company on the datasets are the company who are still in their early bussiness start-up period and trying to develop their products and services.

	FUNDING_STAGE	FUNDING_STAGE_COUNT
1	Series A	951
2	Seed	850
3	Series B	648
4	Series C	463
5	Series D	247
6	Series E	135
7	Unknown	64
8	Series F	62
9	Pre-Seed	58
10	Series G	26
11	Private Equity	21
12	Angel	14
13	Debt Financing	12
14	ICO	5
15	Series H	4
16	Crowdfunding	3

5. **Can you prove how the stage funding can affect the average of funding the company get?**

   To prove that we're gonna find the average funding in each funding_stage column. In the query below we use the Length function because the data type using the varchar. Disclaimer, this query will execute the average length of the character in funding_stage column.

```
SELECT FUNDING_STAGE, AVG(LENGTH(FUNDING_AMOUNT_USD)) AS Average_Funding_Amount
FROM MYTABLE
GROUP BY FUNDING_STAGE
ORDER BY Average_Funding_Amount desc;
```

   The funding stage works by step by step for the company to expand their business and develop their products. The early stage in stage funding is the seed funding where the company start their business step in product development and market research and usually got funding around $500,000 and $2 million. The stage funding can still continue to stage funding stage series H where the company is on the advanced stage of funding. At this stage, the company has typically achieved significant growth and has already gone through multiple funding rounds, including Series A, Series B, Series C, Series D, Series E, Series F, and Series G. So it's not surprising that they get more funding than other funding stages

	FUNDING_STAGE	AVERAGE_FUNDING_AMOUNT
1	Series H	9.25
2	Series G	8.884615384615384615384615384615384615 38
3	Series F	8.8548387096774193548387096774193548 3871
4	Series E	8.725925925925925925925925925925925 92593
5	Private Equity	8.714285714285714285714285714285714285 71
6	Series D	8.558704453441295546558704453441295546 56
7	Series C	8.298056155507559395248380129589632829 37
8	Series B	8.020061728395061728395061728395061728 4
9	Funding	8
10	Debt Financing	8
11	Unkown	8
12	Growth	8
13	Initial Coin Offering	8
14	Unknown	7.671875
15	Series A	7.625657202944269190325972660357518401 68
16	ICO	7.6
17	Crowdfunding	7
18	Grant	7
19	Undisclosed	7
20	Seed	6.903529411764705882352941176470588235 29
21	Angel	6.714285714285714285714285714285714285 71
22	Pre-Seed	6.482758620689655172413793103448275862 07

So it’s being prove that the funding stage can affect the average of funding based on the result we get.