

Исследование данных об инвестициях венчурных фондов в компании-стартапы

Описание проекта:

Работа в написании SQL запросов с базой данных, которая хранит информацию о венчурных фондах и инвестициях в компании-стартапы для дальнейшего их сравнения и анализа.

Данные таблиц:

Теперь можно познакомиться с данными, которые хранят таблицы.

acquisition

Содержит информацию о покупках одних компаний другими.

Таблица включает такие поля:

первичный ключ `id` — идентификатор или уникальный номер покупки;
внешний ключ `acquiring_company_id` — ссылается на таблицу `company` — идентификатор компании-покупателя, то есть той, что покупает другую компанию;

внешний ключ `acquired_company_id` — ссылается на таблицу `company` — идентификатор компании, которую покупают;

`term_code` — способ оплаты сделки:

`cash` — наличными;

`stock` — акциями компании;

`cash_and_stock` — смешанный тип оплаты: наличные и акции.

`price_amount` — сумма покупки в долларах;

`acquired_at` — дата совершения сделки;

`created_at` — дата и время создания записи в таблице;

`updated_at` — дата и время обновления записи в таблице.

company

Содержит информацию о компаниях-стартапах.

первичный ключ `id` — идентификатор, или уникальный номер компании;

`name` — название компании;

`category_code` — категория деятельности компании, например:

`news` — специализируется на работе с новостями;

`social` — специализируется на социальной работе.

`status` — статус компании:

`acquired` — приобретена;
`operating` — действует;
`ipo` — вышла на IPO;
`closed` — перестала существовать.
`founded_at` — дата основания компании;
`closed_at` — дата закрытия компании, которую указывают в том случае, если компании больше не существует;
`domain` — домен сайта компании;
`twitter_username` — название профиля компании в твиттере;
`country_code` — код страны, например, `USA` для США, `GBR` для Великобритании;
`investment_rounds` — число раундов, в которых компания участвовала как инвестор;
`funding_rounds` — число раундов, в которых компания привлекала инвестиции;
`funding_total` — сумма привлечённых инвестиций в долларах;
`milestones` — количество важных этапов в истории компании;
`created_at` — дата и время создания записи в таблице;
`updated_at` — дата и время обновления записи в таблице.

education

Хранит информацию об уровне образования сотрудников компаний.

первичный ключ `id` — уникальный номер записи с информацией об образовании;
внешний ключ `person_id` — ссылается на таблицу `people` — идентификатор человека, информация о котором представлена в записи;
`degree_type` — учебная степень, например:
 `BA` — Bachelor of Arts — бакалавр гуманитарных наук;
 `MS` — Master of Science — магистр естественных наук.
`institution` — учебное заведение, название университета;
`graduated_at` — дата завершения обучения, выпуска;
`created_at` — дата и время создания записи в таблице;
`updated_at` — дата и время обновления записи в таблице.

fund

Хранит информацию о венчурных фондах.

первичный ключ `id` — уникальный номер венчурного фонда;

`name` — название венчурного фонда;
`founded_at` — дата основания фонда;
`domain` — домен сайта фонда;
`twitter_username` — профиль фонда в твиттере;
`country_code` — код страны фонда;
`investment_rounds` — число инвестиционных раундов, в которых фонд принимал участие;
`invested_companies` — число компаний, в которые инвестировал фонд;
`milestones` — количество важных этапов в истории фонда;
`created_at` — дата и время создания записи в таблице;
`updated_at` — дата и время обновления записи в таблице.

`funding_round`

Содержит информацию о раундах инвестиций.

первичный ключ `id` — уникальный номер инвестиционного раунда;
внешний ключ `company_id` — ссылается на таблицу `company` — уникальный номер компании, участвовавшей в инвестиционном раунде;
`funded_at` — дата проведения раунда;
`funding_round_type` — тип инвестиционного раунда, например:
 `venture` — венчурный раунд;
 `angel` — ангельский раунд;
 `series_a` — раунд А.
`raised_amount` — сумма инвестиций, которую привлекла компания в этом раунде в долларах;
`pre_money_valuation` — предварительная, проведённая до инвестиций оценка стоимости компании в долларах;
`participants` — количество участников инвестиционного раунда;
`is_first_round` — является ли этот раунд первым для компании;
`is_last_round` — является ли этот раунд последним для компании;
`created_at` — дата и время создания записи в таблице;
`updated_at` — дата и время обновления записи в таблице.

`investment`

Содержит информацию об инвестициях венчурных фондов в компании-стартапы.

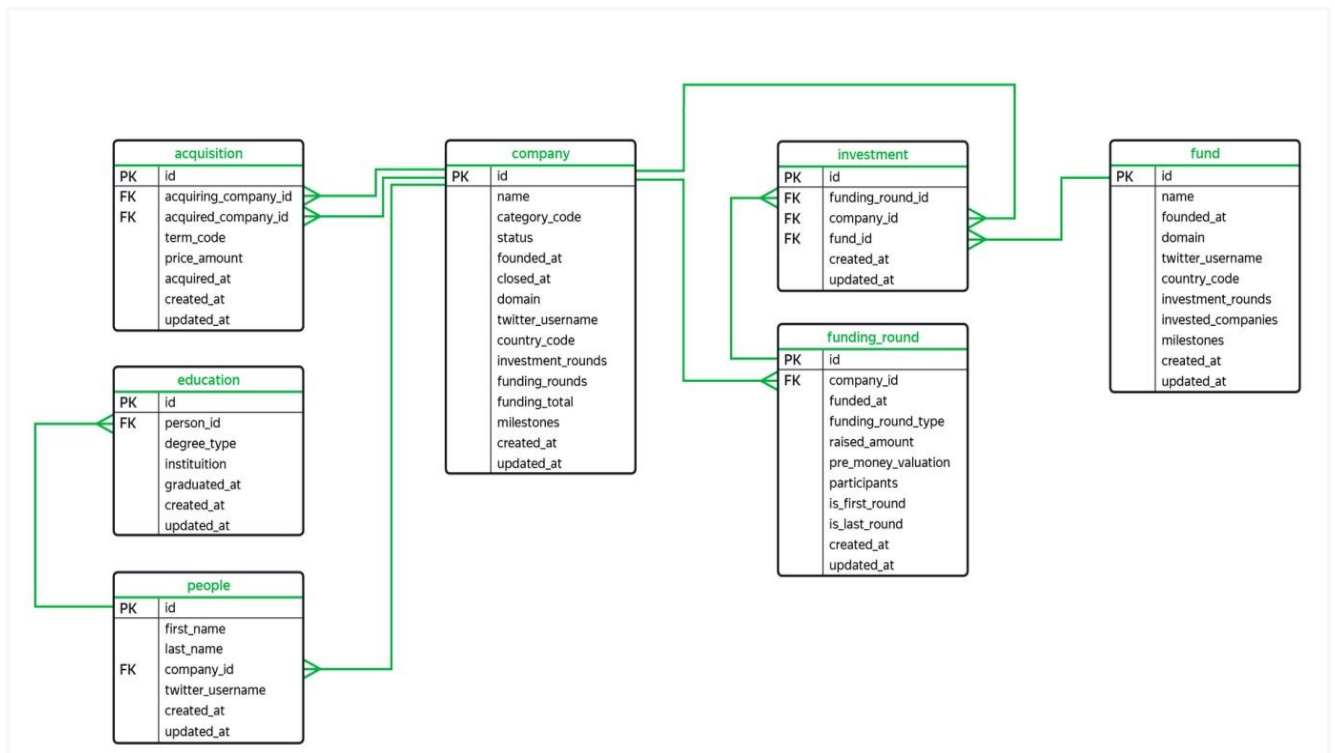
первичный ключ `id` — уникальный номер инвестиции;

внешний ключ `funding_round_id` — ссылается на таблицу `funding_round` — уникальный номер раунда инвестиции;
внешний ключ `company_id` — ссылается на таблицу `company` — уникальный номер компании-стартапа, в которую инвестируют;
внешний ключ `fund_id` — ссылается на таблицу `fund` — уникальный номер фонда, инвестирующего в компанию-стартап;
`created_at` — дата и время создания записи в таблице;
`updated_at` — дата и время обновления записи в таблице.

people

Содержит информацию о сотрудниках компаний-стартапов.

первичный ключ `id` — уникальный номер сотрудника;
`first_name` — имя сотрудника;
`last_name` — фамилия сотрудника;
внешний ключ `company_id` — ссылается на таблицу `company` — уникальный номер компании-стартапа;
`twitter_username` — профиль сотрудника в твиттере;
`created_at` — дата и время создания записи в таблице;
`updated_at` — дата и время обновления записи в таблице.



Задание 1. Количество закрытых компаний

Код

```
1 select count(status)
2 from company
3 where status= 'closed';
```

Результат

count
2584

Задание 2. Количество привлечённых средств для новостных компаний США.Отсортируйте таблицу по убыванию значений в поле `funding_total`.

Код

```
1 select funding_total
2 from company
3 where country_code= 'USA' and category_code= 'news'
4
5 order by funding_total desc;
6
```

Результат

funding_total
6.22553e+08
2.5e+08
1.605e+08
1.28e+08
1.265e+08
7e+07
6.9e+07
6.11322e+07
5.64649e+07

Задание 3. Общая сумма сделок по покупке одних компаний другими в долларах. Отберите сделки, которые осуществлялись только за наличные с 2011 по 2013 год включительно.

Код

```
1 select sum(price_amount)
2 from acquisition
3 where extract(year from cast(acquired_at as date)) between 2011 and 2013
4 and term_code='cash';
```

Результат

sum
1.37762e+11

Задание 4. Имя, фамилия и названия аккаунтов людей в твиттере, у которых названия аккаунтов начинаются на 'Silver'.

Код

```
1 select first_name,last_name,twitter_username
2 from people
3 where twitter_username like 'Silver%';
```

Результат

first_name	last_name	twitter_username
Rebecca	Silver	SilverRebecca
Silver	Teede	SilverMatrixx
Mattias	Guilotte	Silverreven

Задание 5. Выведите на экран всю информацию о людях, у которых названия аккаунтов в твиттере содержат подстроку 'money', а фамилия начинается на 'K'.

Код

```
1 select*
2 from people
3 where twitter_username like '%money%' and last_name like 'K%';
```

Результат

id	first_name	last_name	company_id	twitter_username	created_at
63081	Gregory	Kim		gmoney75	2010-07-13 03:46:28

Задание 6. Для каждой страны отобразите общую сумму привлечённых инвестиций, которые получили компании, зарегистрированные в этой стране. Страну, в которой зарегистрирована компания, можно определить по коду страны. Отсортируйте данные по убыванию суммы.

Код

```
1 select country_code, sum(funding_total) as total
2 from company
3 group by country_code
4 order by total desc;
```

Результат

country_code	total
USA	3.10588e+11
GBR	1.77056e+10
	1.08559e+10
CHN	1.06897e+10
CAN	9.86636e+09
IND	6.14141e+09
DEU	5.76577e+09
FRA	4.59514e+09
ISR	4.48009e+09

Задание 7. Составьте таблицу, в которую войдёт дата проведения раунда, а также минимальное и максимальное значения суммы инвестиций, привлечённых в эту дату.

Оставьте в итоговой таблице только те записи, в которых минимальное значение суммы инвестиций не равно нулю и не равно максимальному значению.

Код

```
1 select funded_at, min(raised_amount),max(raised_amount)
2 from funding_round
3 group by funded_at
4 having min(raised_amount)!=0 and min(raised_amount)!=max(raised_amount);
5
```

Результат

funded_at	min	max
2012-08-22	40000	7.5e+07
2010-07-25	3.27825e+06	9e+06
2002-03-01	2.84418e+06	8.95915e+06
2010-10-11	28000	2e+08
2007-01-18	5.5e+06	2.3e+07
2007-02-27	1.29e+06	3.6e+07
2006-01-05	8.9e+06	2.65e+07
2011-10-31	35000	2.5e+07
2012-10-27	500000	9.3e+06

Задание 8. Создайте поле с категориями:

Для фондов, которые инвестируют в 100 и более компаний, назначьте категорию **high_activity**.

Для фондов, которые инвестируют в 20 и более компаний до 100, назначьте категорию **middle_activity**.

Если количество инвестируемых компаний фонда не достигает 20, назначьте категорию **low_activity**.

Отобразите все поля таблицы **fund** и новое поле с категориями.

Код

```
1 select *,
2 CASE
3   WHEN invested_companies >=100 THEN 'high_activity'
4   WHEN invested_companies >=20 and invested_companies<100 THEN 'middle_activity'
5   WHEN invested_companies <20 THEN 'low_activity'
6 END
7 FROM fund;
8
```

Результат

id	name	founded_at	domain	twitter_username	country_code	investment_rounds	invested_companies	milestones	created_at	updated_at	case
13131						0	0	0	2013-08-19 18:46:55	2013-08-19 19:55:07	low_activity
1	Greylock Partners	1965-01-01	greylock.com	greylockvc	USA	307	196	0	2007-05-25 20:18:23	2012-12-27 00:42:24	high_activity
10	Mission Ventures	1996-01-01	missionventures.com		USA	58	33	0	2007-06-05 05:24:58	2013-10-10 22:06:31	middle_activity
100	Kapor Enterprises, Inc.		kei.com		USA	2	1	0	2007-07-12 09:42:21	2008-11-21 05:41:53	low_activity
1000	Speed Ventures					0	0	1	2008-04-13 23:52:27	2008-12-10 09:37:18	low_activity
10000	3x5 Special Opportunity Partners					4	4	0	2012-10-26 03:16:38	2012-10-26 03:16:38	low_activity

Задание 9. Для каждой из категорий, назначенных в предыдущем задании, посчитайте округлённое до ближайшего целого числа среднее количество инвестиционных раундов, в которых фонд принимал участие. Выведите на экран категории и среднее число инвестиционных раундов. Отсортируйте таблицу по возрастанию среднего.

Код

```
1 SELECT
2     CASE
3         WHEN invested_companies >= 100 THEN 'high_activity'
4         WHEN invested_companies >= 20 THEN 'middle_activity'
5         ELSE 'low_activity'
6     END AS activity, round(avg(investment_rounds)) AS rounded
7 FROM fund
8 GROUP BY 1
9 ORDER BY 2;
10
```

Результат

activity	rounded
low_activity	2
middle_activity	51
high_activity	252

Задание 10. Проанализируйте, в каких странах находятся фонды, которые чаще всего инвестируют в стартапы.

Для каждой страны посчитайте минимальное, максимальное и среднее число компаний, в которые инвестировали фонды этой страны, основанные с 2010 по 2012 год включительно. Исключите страны с фондами, у которых минимальное число компаний, получивших инвестиции, равно нулю.

Выгрузите десять самых активных стран-инвесторов: отсортируйте таблицу по среднему количеству компаний от большего к меньшему. Затем добавьте сортировку по коду страны в лексикографическом порядке.

Код

```
1 select
  country_code,min(invested_companies),max(invested_companies),avg(invested_companies)
2 from fund
3 where extract(year from cast(founded_at as date)) between 2010 and 2012
4 group by country_code
5 having min(invested_companies)!=0
6 order by avg(invested_companies) desc
7 limit 10;
8
```

Результат

country_code	min	max	avg
BGR	25	35	30
CHL	29	29	29
UKR	8	10	9
LTU	5	5	5
IRL	4	5	4.5

Задание 11. Отобразите имя и фамилию всех сотрудников стартапов. Добавьте поле с названием учебного заведения, которое окончил сотрудник, если эта информация известна.

Код

```
1 select first_name,last_name, institution
2 from people as p
3 LEFT JOIN education as e on p.id= e.person_id;
```

Результат

first_name	last_name	institution
John	Green	Washington University, St. Louis
John	Green	Boston University
David	Peters	Rice University
Dan	Birdwhistell	University of Cambridge
Gal	Cohen	Tel Aviv University
Chris	Treadaway	University of Texas
Chris	Treadaway	Louisiana State University
Sam	Lessin	Harvard University
Guy	Levy-Yurista	University of Pennsylvania - The Wharton School

Задание 12. Для каждой компании найдите количество учебных заведений, которые окончили её сотрудники. Выведите название компании и число уникальных названий учебных заведений. Составьте топ-5 компаний по количеству университетов.

Код

```
1 SELECT c.name,  
2 count(distinct ed.institution) as quantity  
3 from company as c  
4 INNER JOIN people as p on c.id=p.company_id  
5 INNER JOIN education as ed on p.id=ed.person_id  
6 group by c.name  
7 order by quantity desc  
8 limit 5;
```

Результат

name	quantity
Google	167
Yahoo!	115
Microsoft	111
Knight Foundation	74
Comcast	66

Задание 13. Составьте список с уникальными названиями закрытых компаний, для которых первый раунд финансирования оказался последним.

Код

```
1 select distinct c.name
2 from company as c
3 INNER JOIN funding_round as fr on c.id=fr.company_id
4 WHERE c.status='closed' and fr.is_first_round=1
5 and fr.is_last_round=1;
```

Результат

name
10BestThings
11i Solutions
169 ST.
1bib
1Cast
1DayMakeover
25eight
27 Perry
2Win-Solutions

Задание 14. Составьте список уникальных номеров сотрудников, которые работают в компаниях, отображенных в предыдущем задании.

Код

```
1 SELECT distinct people.id
2 FROM people
3 JOIN company ON people.company_id = company.id
4 JOIN funding_round ON company.id = funding_round.company_id
5 WHERE is_first_round = 1 AND is_last_round = 1
6     AND status = 'closed';
```

Результат

id
62
97
98
225
226
227
281
282
283

Задание 15. Составьте таблицу, куда войдут уникальные пары с номерами сотрудников из предыдущей задачи и учебным заведением, которое окончил сотрудник.

Код

```
1 select distinct(p.id), ed.institution
2 FROM people as p
3 JOIN company as c ON p.company_id = c.id
4 JOIN funding_round as fr ON c.id = fr.company_id
5 JOIN education as ed on ed.person_id=p.id
6 WHERE fr.is_first_round = 1 AND fr.is_last_round = 1
7 AND c.status = 'closed';
```

Результат

id	institution
349	AKI
349	ArtEZ Hogeschool voor de Kunsten
349	Rijks Akademie
699	Imperial College
779	Harvard University
779	Stanford University
968	University of Notre Dame
972	The University of Texas at Austin
1107	CDI, Sydney

Задание 16. Посчитайте количество учебных заведений для каждого сотрудника из предыдущего задания. При подсчёте учитывайте, что некоторые сотрудники могли окончить одно и то же заведение дважды.

Код

```
1 SELECT p.id, COUNT(distinct e.id) as educ
2 FROM company c
3     JOIN funding_round f ON c.id = f.company_id
4     JOIN people p ON c.id = p.company_id
5     JOIN education e ON e.person_id = p.id
6 WHERE c.status = 'closed'
7     AND f.is_first_round = 1
8     AND f.is_last_round = 1
9 GROUP BY p.id
```

Результат

id	educ
349	3
699	1
779	2
968	1
972	1
1107	1
1444	2
1833	1
1911	1

Задание 17. Дополните предыдущий запрос и выведите среднее число учебных заведений (всех, не только уникальных), которые окончили сотрудники разных компаний. Нужно вывести только одну запись, группировка здесь не понадобится.

Код

```
1 SELECT AVG(new.educ)
2 FROM (SELECT DISTINCT p.id, COUNT(DISTINCT e.id) as educ
3 FROM company c
4     JOIN funding_round f ON c.id = f.company_id
5     JOIN people p ON c.id = p.company_id
6     JOIN education e ON e.person_id = p.id
7 WHERE c.status = 'closed'
8     AND f.is_first_round = 1
9     AND f.is_last_round = 1
10 GROUP BY p.id) as new;
11
```

Результат

avg

1.41509

Задание 18. Напишите похожий запрос: выведите среднее число учебных заведений (всех, не только уникальных), которые окончили сотрудники Facebook*.

Код

```
1 SELECT AVG(new.quantity)
2 FROM
3 (SELECT p.id, count(ed.institution) as quantity
4 from company as c
5 JOIN people as p on p.company_id=c.id
6 JOIN education as ed on ed.person_id=p.id
7 WHERE c.name='Facebook'
8 GROUP BY p.id) AS new;
```

Результат

avg

1.51111

Задание 19. Составьте таблицу из полей:

name_of_fund — название фонда;

name_of_company — название компании;

amount — сумма инвестиций, которую привлекла компания в раунде.

В таблицу войдут данные о компаниях, в истории которых было больше шести важных этапов, а раунды финансирования проходили с 2012 по 2013 год включительно.

Код

```
1 SELECT f.name as name_of_fund,  
2 c.name as name_of_company, fr.raised_amount as amount  
3 from investment as i  
4 JOIN company as c on c.id=i.company_id  
5 JOIN fund as f on f.id=i.fund_id  
6 JOIN funding_round as fr on i.funding_round_id=fr.id  
7 WHERE c.milestones>6  
8 AND CAST(fr.funded_at as date) BETWEEN '01.01.2012' AND '2013-12-31';
```

Результат

name_of_fund	name_of_company	amount
Advance Publication	Gigya	1.53e+07
Mayfield Fund	Gigya	1.53e+07
Benchmark	Gigya	1.53e+07
DAG Ventures	Gigya	1.53e+07
Mitsui Global Investment	OpenX	2.50112e+07
Accel Partners	OpenX	2.50112e+07
Presidio Ventures	OpenX	2.50112e+07
Index Ventures	OpenX	2.50112e+07
Samsung Ventures	OpenX	2.50112e+07

Задание 20. Выгрузите таблицу, в которой будут такие поля:

название компании-покупателя;
сумма сделки;
название компании, которую купили;
сумма инвестиций, вложенных в купленную компанию;
доля, которая отображает, во сколько раз сумма покупки превысила
сумму вложенных в компанию инвестиций, округлённая до
ближайшего целого числа.

Не учитывайте те сделки, в которых сумма покупки равна нулю. Если сумма инвестиций в компанию равна нулю, исключите такую компанию из таблицы. Отсортируйте таблицу по сумме сделки от большей к меньшей, а затем по названию купленной компании в лексикографическом порядке. Ограничьте таблицу первыми десятью записями.

Код

```
1 SELECT c1.name AS name1,  
2        a.price_amount,  
3        c2.name AS name2,  
4        c2.funding_total,  
5        round(a.price_amount / c2.funding_total)  
6 FROM acquisition AS a  
7 LEFT JOIN company AS c1 ON a.acquiring_company_id=c1.id  
8 LEFT JOIN company AS c2 ON a.acquired_company_id=c2.id  
9  
10 WHERE a.price_amount > 0 AND c2.funding_total>0  
11 ORDER BY a.price_amount DESC, name2
```

Результат

name1	price_amount	name2	funding_total	round
Microsoft	8.5e+09	Skype	7.6805e+07	111
Scout Labs	4.9e+09	Varian Semiconductor Equipment Associates	4.8e+06	1021
Broadcom	3.7e+09	Aeluros	7.97e+06	464
Broadcom	3.7e+09	NetLogic Microsystems	1.88527e+08	20
Level 3 Communications	3e+09	Global Crossing	4.1e+07	73
Yahoo!	2.87e+09	GeoCities	4e+07	72

Задание 21. Выгрузите таблицу, в которую войдут названия компаний из категории **social**, получившие финансирование с 2010 по 2013 год включительно. Проверьте, что сумма инвестиций не равна нулю. Выведите также номер месяца, в котором проходил раунд финансирования.

Код

```
1 select c.name, extract(month from cast(fr.funded_at as date))
2 from company as c
3 JOIN funding_round as fr on c.id=fr.company_id
4 WHERE c.category_code='social' and
5 extract(year from cast(fr.funded_at as date)) between 2010 and 2013
6 and fr.raised_amount!=0;
```

Результат

name	date_part
Klout	1
WorkSimple	3
HengZhi	1
Twitter	1
SocialGO	1
ThisNext	1
Tagged	1
LikeMe.Net	2
Busuu	10

Задание 22. Отберите данные по месяцам с 2010 по 2013 год, когда проходили инвестиционные раунды. Сгруппируйте данные по номеру месяца и получите таблицу, в которой будут поля: номер месяца, в котором проходили раунды; количество уникальных названий фондов из США, которые инвестировали в этом месяце; количество компаний, купленных за этот месяц; общая сумма сделок по покупкам в этом месяце.

```
WITH USA AS (SELECT EXTRACT(MONTH FROM fr.funded_at) AS month_number,
                    COUNT(DISTINCT f.name) AS usa_funds
              FROM funding_round AS fr
              LEFT JOIN investment AS inv ON inv.funding_round_id = fr.id
              LEFT JOIN fund AS f ON inv.fund_id = f.id
              WHERE EXTRACT(YEAR FROM fr.funded_at) BETWEEN 2010 AND 2013
                    AND f.country_code = 'USA'
              GROUP BY month_number),
  comp AS (SELECT EXTRACT(MONTH FROM acquired_at) AS month_number,
                    COUNT(acquired_company_id) AS company_number,
                    SUM(price_amount) AS total_sum
            FROM acquisition
            WHERE EXTRACT(YEAR FROM acquired_at) BETWEEN 2010 AND 2013
            GROUP BY month_number)
SELECT USA.month_number,
       USA.usa_funds,
       comp.company_number,
       comp.total_sum
FROM USA JOIN comp ON USA.month_number = comp.month_number;
```

Результат

month_number	usa_funds	company_number	total_sum
1	815	600	2.71083e+10
2	637	418	4.13903e+10
3	695	458	5.95016e+10
4	718	411	3.03837e+10
5	695	532	8.60122e+10
6	785	525	5.20883e+10
7	803	488	4.98541e+10
8	726	454	7.77093e+10
9	793	491	6.97409e+10

Задание 23. Составьте сводную таблицу и выведите среднюю сумму инвестиций для стран, в которых есть стартапы, зарегистрированные в 2011, 2012 и 2013 годах. Данные за каждый год должны быть в отдельном поле. Отсортируйте таблицу по среднему значению инвестиций за 2011 год от большего к меньшему.

Код

```

1 WITH
2 y_2011 as (SELECT country_code,AVG(funding_total) as year_2011
3           FROM company
4           WHERE extract(year from founded_at)=2011
5           GROUP BY country_code),
6
7 y_2012 as (SELECT country_code,AVG(funding_total) as year_2012
8           FROM company
9           WHERE extract(year from founded_at)=2012
10          GROUP BY country_code),
11
12 y_2013 as (SELECT country_code,AVG(funding_total) as year_2013
13           FROM company
14           WHERE extract(year from founded_at)=2013
15           GROUP BY country_code)
16
17 SELECT y_2011.country_code,year_2011,year_2012,year_2013
18 FROM y_2011
19 JOIN y_2012 on y_2011.country_code=y_2012.country_code
20 JOIN y_2013 on y_2013.country_code=y_2012.country_code
21 ORDER BY year_2011 DESC;

```

Результат

country_code	year_2011	year_2012	year_2013
PER	4e+06	41000	25000
USA	2.24396e+06	1.20671e+06	1.09336e+06
HKG	2.18078e+06	226227	0
PHL	1.75e+06	4218.75	2500
ARE	1.718e+06	197222	35333.3
JPN	1.66431e+06	674720	50000
AUT	1.5342e+06	147806	85773.3
BRA	1.38007e+06	240639	67944.4
DEU	1.1288e+06	1.32915e+06	66612.7

