

# SQL Queries for Kickstarter data

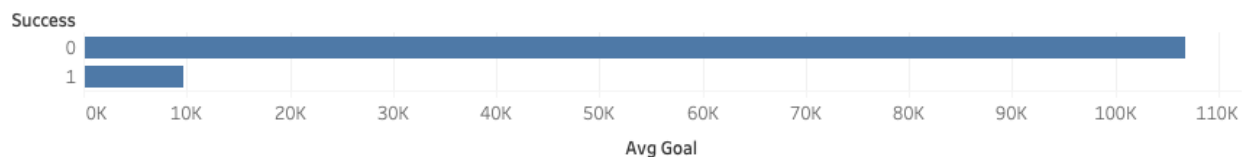
## Part 1- Preliminary Data Analysis

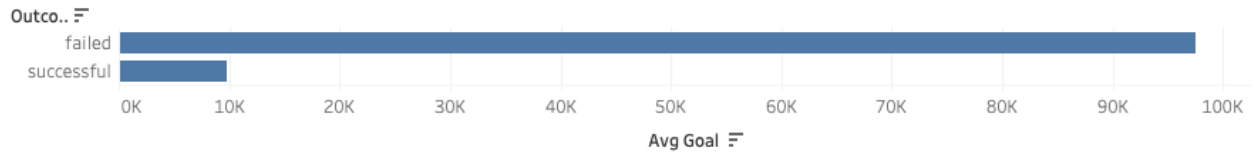
### 1. Goals vs. Success

Are the goals for dollars raised significantly different between campaigns that are successful and unsuccessful?

```
-- Avg_goal_by_campaign_success
select
  case
    when outcome = "successful" then 1
    else 0
    end as Success,
    round(avg(goal),2) as avg_goal
from campaign
group by 1
;
```

```
-- Avg_goal_by_outcome
select
  outcome,
  round(avg(goal),2) as avg_goal
from campaign
group by 1
;
```





Goals in dollars is significantly different between campaigns that are successful and unsuccessful.

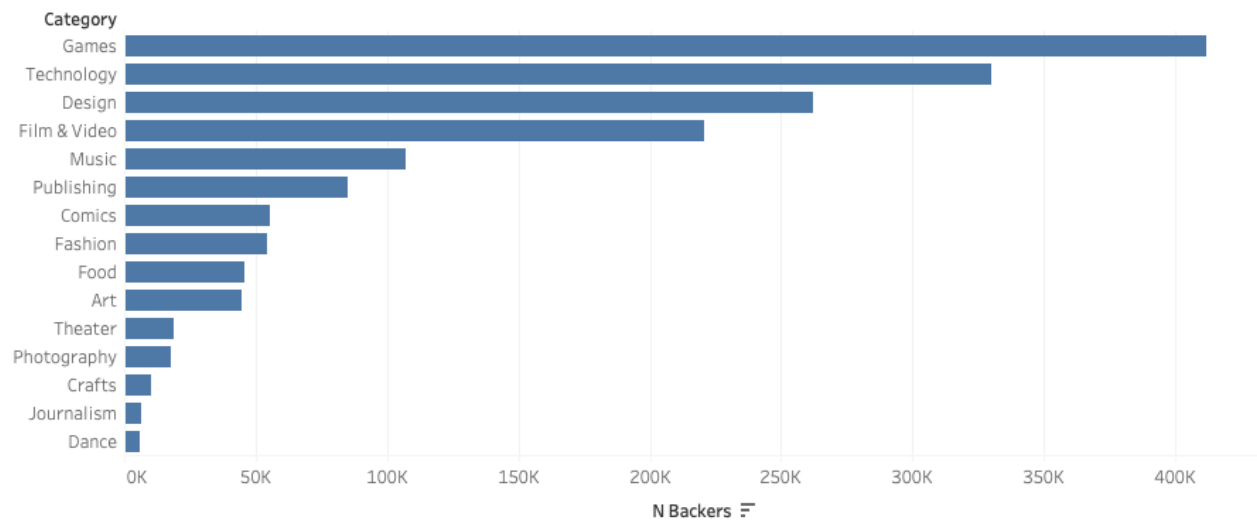
## 2. Categories vs. Backers

**What are the top/bottom 3 categories with the most backers?**

```
with cte as(
select
  a.*,
  b.name as sub_category,
  c.name as category
from campaign a
left join sub_category b on
  a.sub_category_id = b.id
left join category c on
  b.category_id = c.id
)

select
  category,
  sum(backers) as n_backers
from cte
group by 1
order by 2 desc
;
```

## Category by Backers



Sum of N Backers for each Category.

- Top3: Game, Technology, Design
- Bottom3: Crafts, Journalism, Dance

## What are the top/bottom 3 subcategories by backers?

```
with cte as(
select
  a.*,
  b.name as sub_category,
  c.name as category
from campaign a
left join sub_category b on
  a.sub_category_id = b.id
left join category c on
  b.category_id = c.id
)

select
  sub_category,
  sum(backers) as n_backers
from cte
group by 1
order by 2 desc
;
```

### Subcategory by backers



- Top 3: Tabletop Games, Product Design, Video Games
- Bottom 3: Latin, Photo, Glass

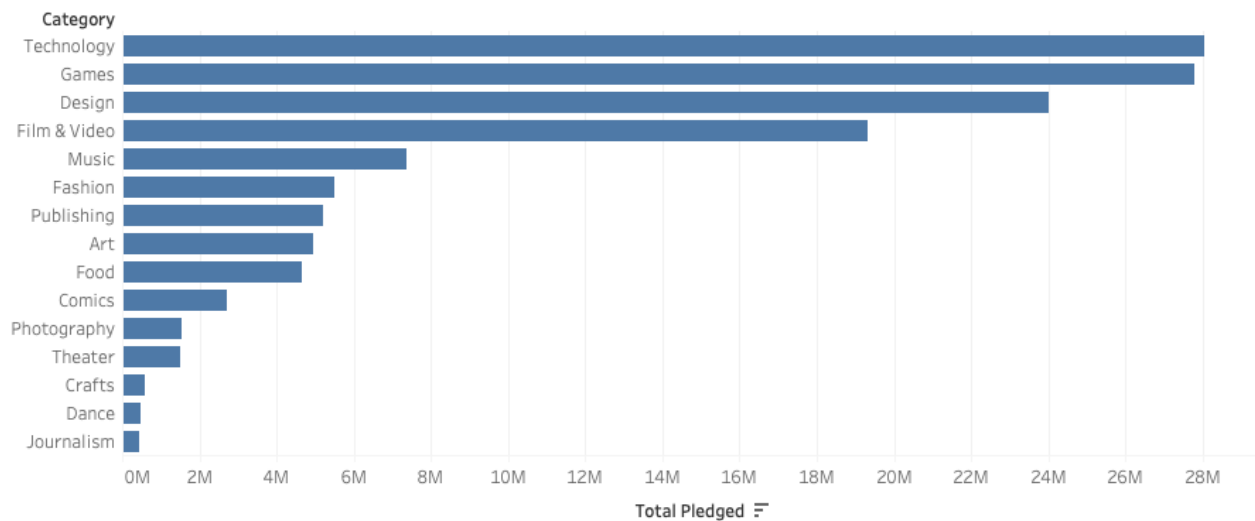
## 3. Categories v.s. Money Raised

**What are the top/bottom 3 categories that have raised the most money?**

```
-- Top/Bottom3_category_by_pledged
with cte as(
select
  a.*,
  b.name as sub_category,
  c.name as category
from campaign a
left join sub_category b on
  a.sub_category_id = b.id
left join category c on
  b.category_id = c.id
)

select
  category,
  round(sum(pledged),2) as total_pledged
from cte
group by 1
order by 2 desc
;
```

## Category by pledged



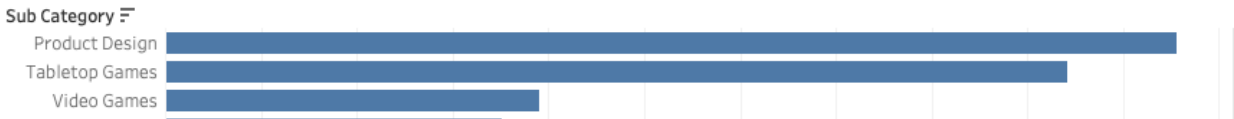
Sum of Total Pledged for each Category.

- Top3: Technology, Game, Design
- Bottom3: Crafts, Dance, Journalism

## What are the top/bottom 3 subcategories that have raised the most money?

```
with cte as(
select
  a.*,
  b.name as sub_category,
  c.name as category
from campaign a
left join sub_category b on
  a.sub_category_id = b.id
left join category c on
  b.category_id = c.id
)

select
  sub_category,
  round(sum(pledged),2) as total_pledged
from cte
group by 1
order by 2 desc
;
```



- Top 3: Product Design, Tabletop Games, Video Games
- Bottom 3: Latin, Crochet, Glass

## 4. Most Successful Board Games

**What was the amount the most successful board game company raised? How many backers did they have?**

Assuming "most successful" is determined by amount of money raised

```
with cte as(
select
  a.*,
  b.name as sub_category,
  c.name as category
from campaign a
left join sub_category b on
  a.sub_category_id = b.id
left join category c on
  b.category_id = c.id
)

select `name`, goal, pledged, backers
from cte
where outcome = 'successful' and
  sub_category = 'Tabletop Games'
order by pledged desc
limit 3
;
```

name	goal	pledged	backers
Gloomhaven (Second Printing)	100000	3999800	40642
Ghostbusters": The Board Game	250000	1546270	8396
Shadows of Brimstone	30000	1341300	4727

The most successful board game ever raised is **Gloomhaven (Second Printing)**, raised around **40** times the goal. They have 40642 backers.

## 5. Country v.s. Campaigns

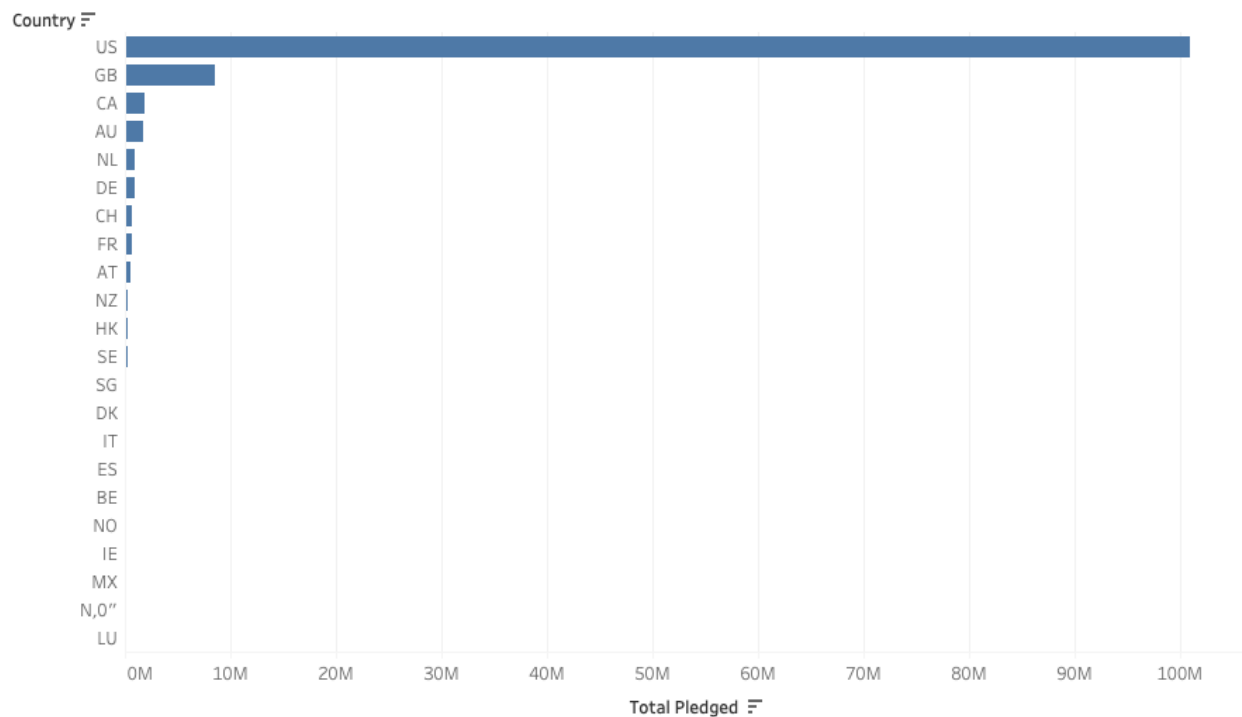
Rank the top three countries with the most successful campaigns:

### 1. in terms of dollars (total amount pledged)

```
-- Top3_country_by_dollars
with cte as(
select
  a.*,
  b.name as sub_category,
  c.name as category,
  d.name as country
from campaign a
left join sub_category b on
  a.sub_category_id = b.id
left join category c on
  b.category_id = c.id
left join country d on
  a.country_id = d.id
)
, rnk_pledged as(
select
  country,
  round(sum(pledged),2) as total_pledged,
  rank()over(order by sum(pledged) desc) as rnk
from cte
where outcome = 'successful'
group by 1
order by 2 desc
)
select *
from rnk_pledged
```

```
where rnk <=3  
;
```

## Pledged by Country



## 2. in terms of the number of campaigns backed.

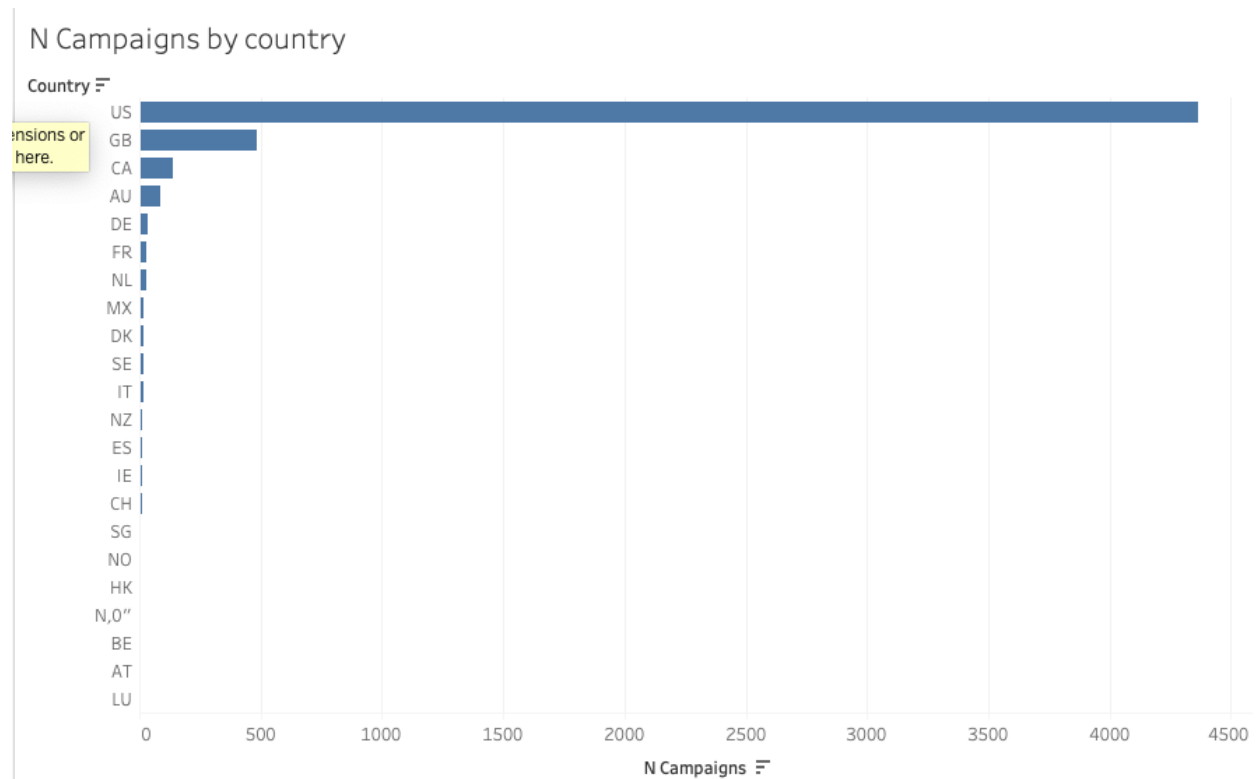
```
-- Top3_country_by_Ncampaigns  
with cte as(  
select  
    a.*,  
    b.name as sub_category,  
    c.name as category,  
    d.name as country  
from campaign a  
left join sub_category b on  
    a.sub_category_id = b.id  
left join category c on  
    b.category_id = c.id  
left join country d on  
    a.country_id = d.id  
)
```



```

select
  country,
  count(distinct id) as n_campaigns
from cte
where outcome = 'successful'
group by 1
order by 2 desc
;

```



## 6. Campaign Duration v.s. Pledged

Do longer, or shorter campaigns tend to raise more money? Why?

```

-- Since functions like datediff,
-- timestampdiff can't be used in this interface,

```

```

-- the below approximate the campaign duration in days
with cte as(
select
    id, name,
    date(launched) as start, date(deadline) as end,
    365*(strftime('%Y', deadline) - strftime('%Y', launched)) +
    30*(strftime('%m', deadline) - strftime('%m', launched)) +
    strftime('%d', deadline) - strftime('%d', launched) as duration,
    pledged
from campaign
where outcome = 'successful'
)
select
    case
        when duration <=30 then '1 month'
        when duration <=60 then '2 months'
        else '3 months'
    end as length,
    round(avg(pledged),2) as avg_pledged
from cte
group by 1
;

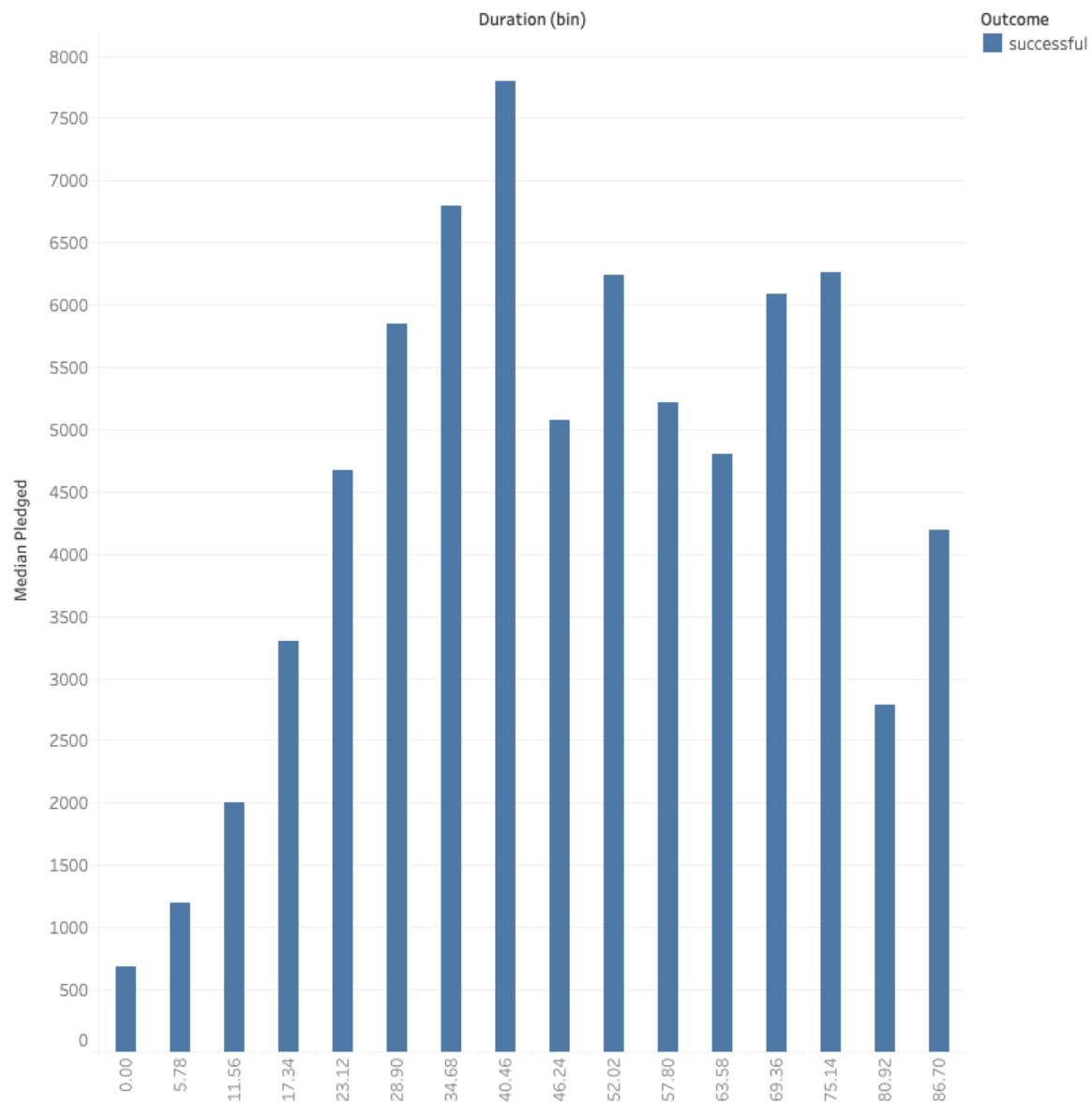
-- using datediff in mysql
with campaign_length as(
select
    *,
    cast(datediff(deadline, launched) as char) as duration
from campaign
),
campaign_length2 as(
select
    *,
    case
        when duration <= 31 then "1 month"
        when duration <= 62 then "2 months"
        else "3 months"
    end as duration_segment
from campaign_length
)
select
    duration_segment,
    round(avg(goal),2) as avg_goals,
    round(avg(pledged),2) as avg_raised,
    round(avg(backers)) as avg_backers,
    round(avg(pledged/backers), 2) as cost_per_backers,
    avg(case when outcome = 'successful' then 1 else 0 end ) as success_rate,
    count(distinct id) as n_campaign
from campaign_length2
where
    -- (outcome = 'successful' or outcome = 'failed') and
    -- sub_category_id = 14 and

```

```
-- goal >= 15000 and goal <=40000
group by 1
order by 1
;
```

It seems like average money raised peaked at duration of around 40 days overall. However, the actual relationship between the two variables needs to be further examined.

Money raised peaked at around 40 days



Median of Pledged for each Duration (bin). Color shows details about Outcome. The data is filtered on Exclusions (Duration,Outcome), which keeps 308 members. The view is filtered on Outcome, which keeps successful.