Assignment: Joining and Analysis Project

1. Which quiz generated the most views? What type of quiz is it?

```
select  a.quiz_id, sum(num_views) as total_views, quiz_type
FROM daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
group by a.quiz_id, quiz_type
order by total_views desc
```

The quiz type that generated the most views was personality type quizzes.

2. Which quiz has the highest completion rate? What type of quiz is it and when was it published?

```
select  a.quiz_id, (sum(num_completes) /  sum(num_views) ) AS completion_rate ,
quiz_type, published
FROM daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
group by a.quiz_id, quiz_type, published
order by completion_rate desc;
```

Most completed types were personality types. It was published on 1552243095. It Had a completion rate of 5.12469032384551, this may be garbage data since that's an inaccurate percentage.

So the highest completion rate that was below 1 was quiz id 2383572 which was a personality type quiz published on 1549400159

3. Which type of quiz generated the most views? What about on average, which type of quiz generates the most views?

```
SELECT a.quiz_id, b.quiz_type, sum(a.num_views) as num_of_views, avg(a.num_views)
as Average_views
from daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
group by a.quiz_id, b.quiz_type
order by num_of_views desc
```

Personality type quizzes generated the most views, the same could be said for average views.

4. Which type of quiz has the highest completion rate?

```
select  a.quiz_id, (sum(num_completes) /  sum(num_views) ) AS completion_rate , quiz_type
FROM daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
group by quiz_type, a.quiz_id
order by completion_rate desc
```

Personality type quizzes had the highest completion rate

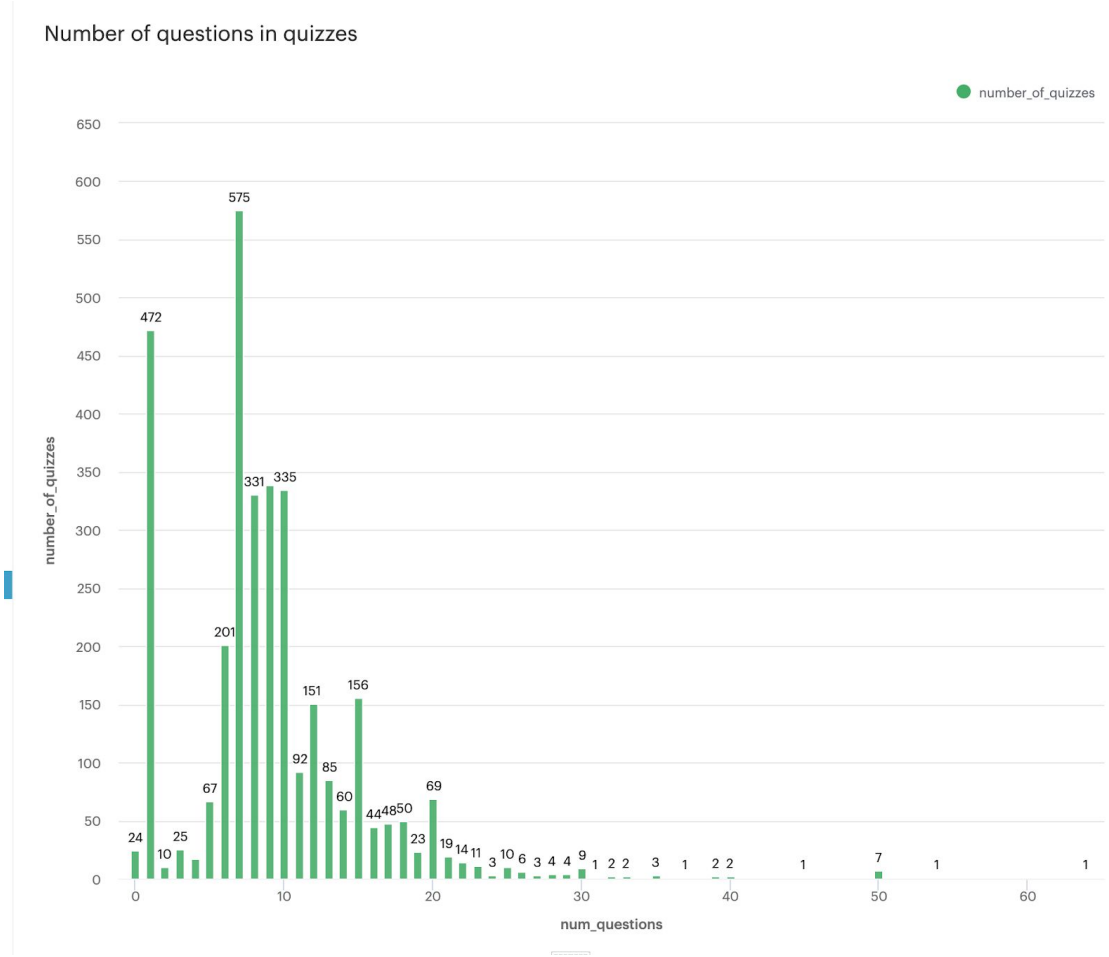
5. Which type of quiz had the highest share rate?

```
select  a.quiz_id, (sum(num_shares) /  sum(num_views) ) AS share_rate , quiz_type
FROM daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
group by quiz_type, a.quiz_id
order by share_rate desc
```

Personality had the highest share rate

6. Create a histogram chart where the x-axis is the number of questions and the y-axis is the number of quizzes.

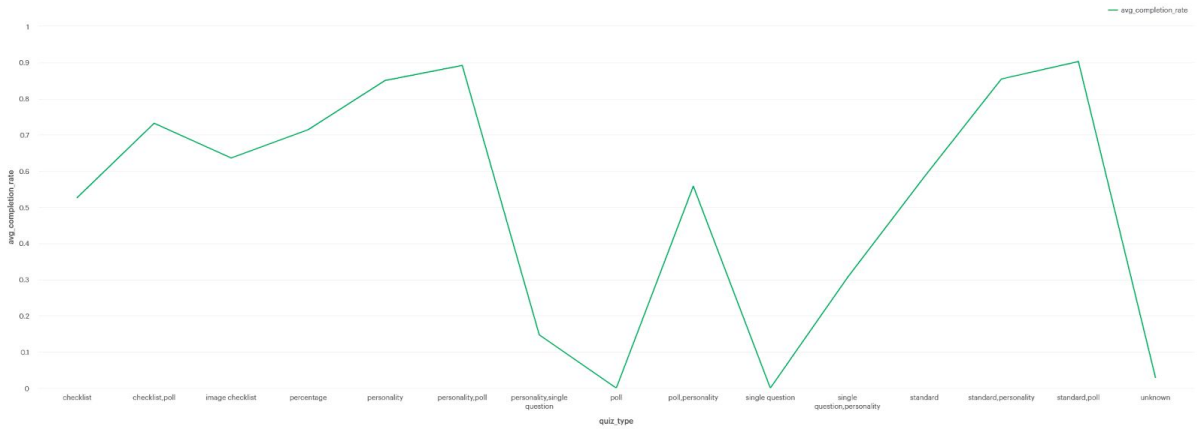
```
select count(num_questions) as number_of_quizzes, num_questions
FROM quiz_metadata
group by  num_questions
order by number_of_quizzes DESC
```



Analysis Project

What types of quizzes perform best?

```
select count(quiz_type), quiz_type,(avg(num_completes) /  avg(num_views) ) AS
avg_completion_rate
FROM daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
group by quiz_type
having (avg(num_completes) /  avg(num_views) ) < 1
order by avg_completion_rate desc
```



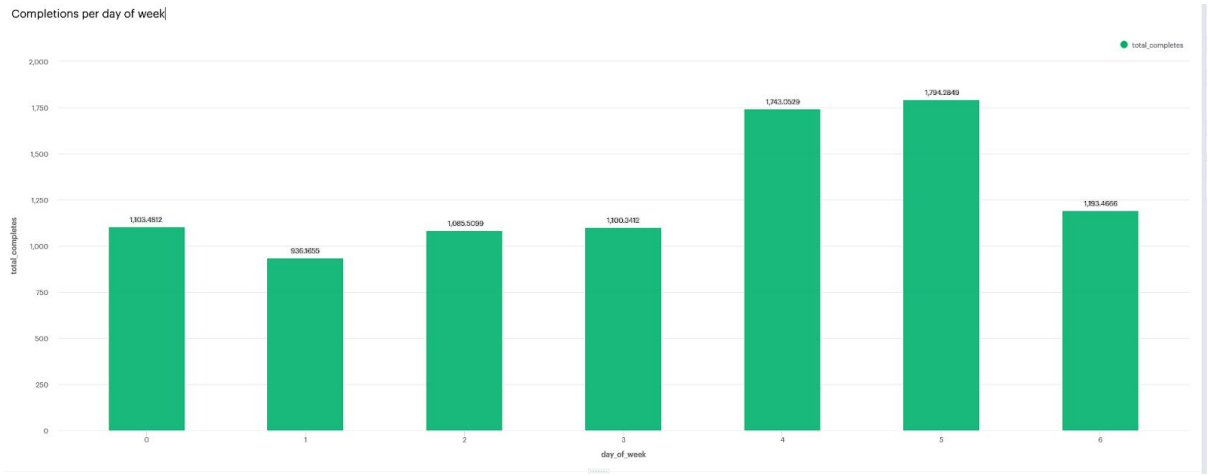
	count	quiz_type	avg_completion_rate
1	68	standard,poll	0.9026239849674519
2	102	personality,poll	0.8917492890156519
3	23	standard,personality	0.8544423925310352
4	64866	personality	0.8505198939504879
5	2	checklist,poll	0.7322033898305085
6	763	percentage	0.7144122897556023
7	2265	image checklist	0.6360014836968625
8	20745	standard	0.5840978335475204
9	22	poll,personality	0.558083489910955
10	4150	checklist	0.5264546099748371
11	66	single question,personality	0.3050230623616925
12	33	personality,single question	0.1469739354380356
13	675	unknown	0.029555673418290067
14	648	single question	0
15	769	poll	0

The graph above shows the average completion rate for each type of quiz. Personality polls of various types show to be performing best. Single poll questions show to be performing the worst.

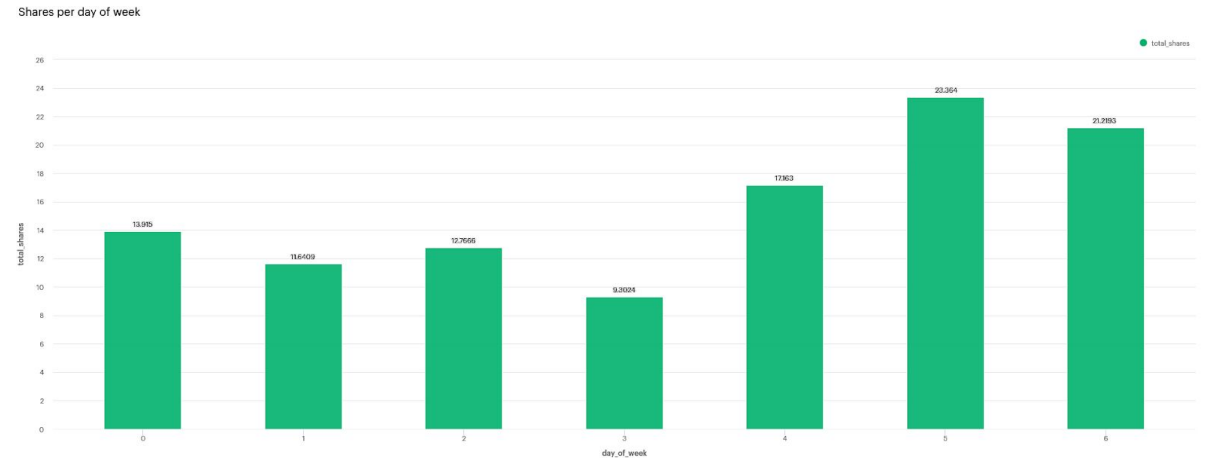
When should we publish a quiz?

```
SELECT extract(dow from date(to_timestamp("published"))) as day_of_week ,
avg(num_views) as total_views, avg(num_completes) as total_completes,
avg(num_shares) as total_shares
from daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
group by day_of_week
order by total_views, total_completes, total_shares DESC
```

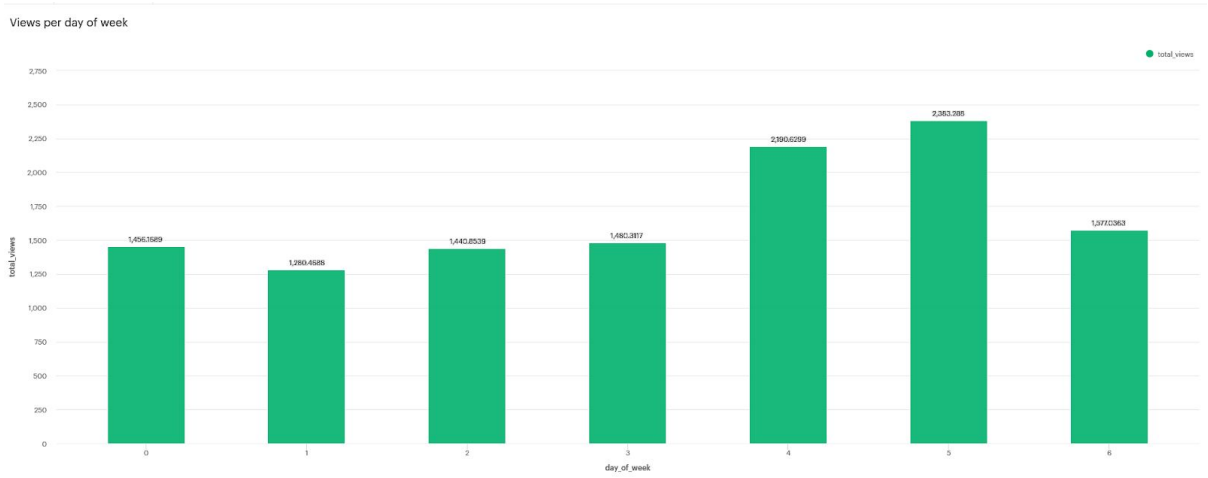
Quiz completions per days of week



Quiz shares per days of week



Quiz views per days of week



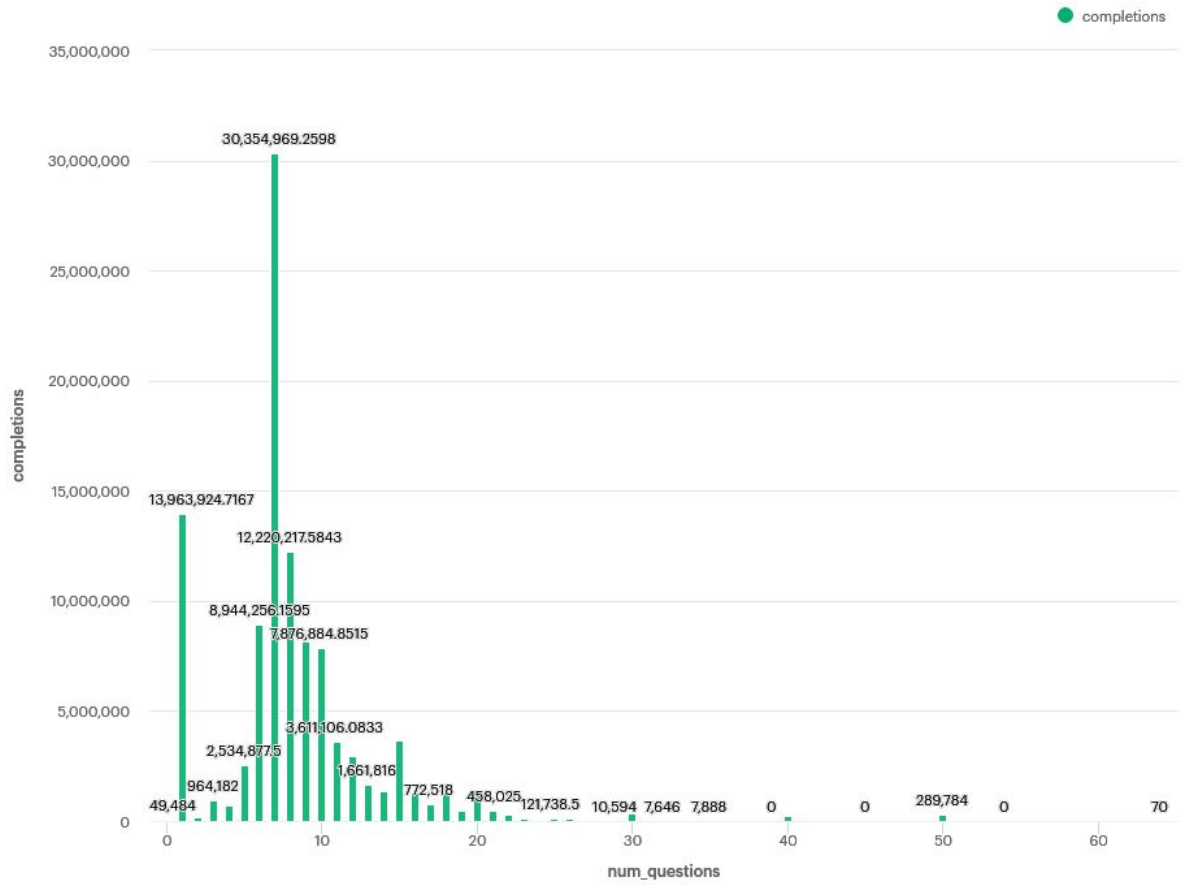
(Graphs 0 - 6, 0 represents Sunday, 1= Monday etc)

The graphs show they generally have more views, complets and shares during the end of the working week, so thursday Friday. So recommendation is to publish a quiz on Thursday to get the two most busiest days of the week.

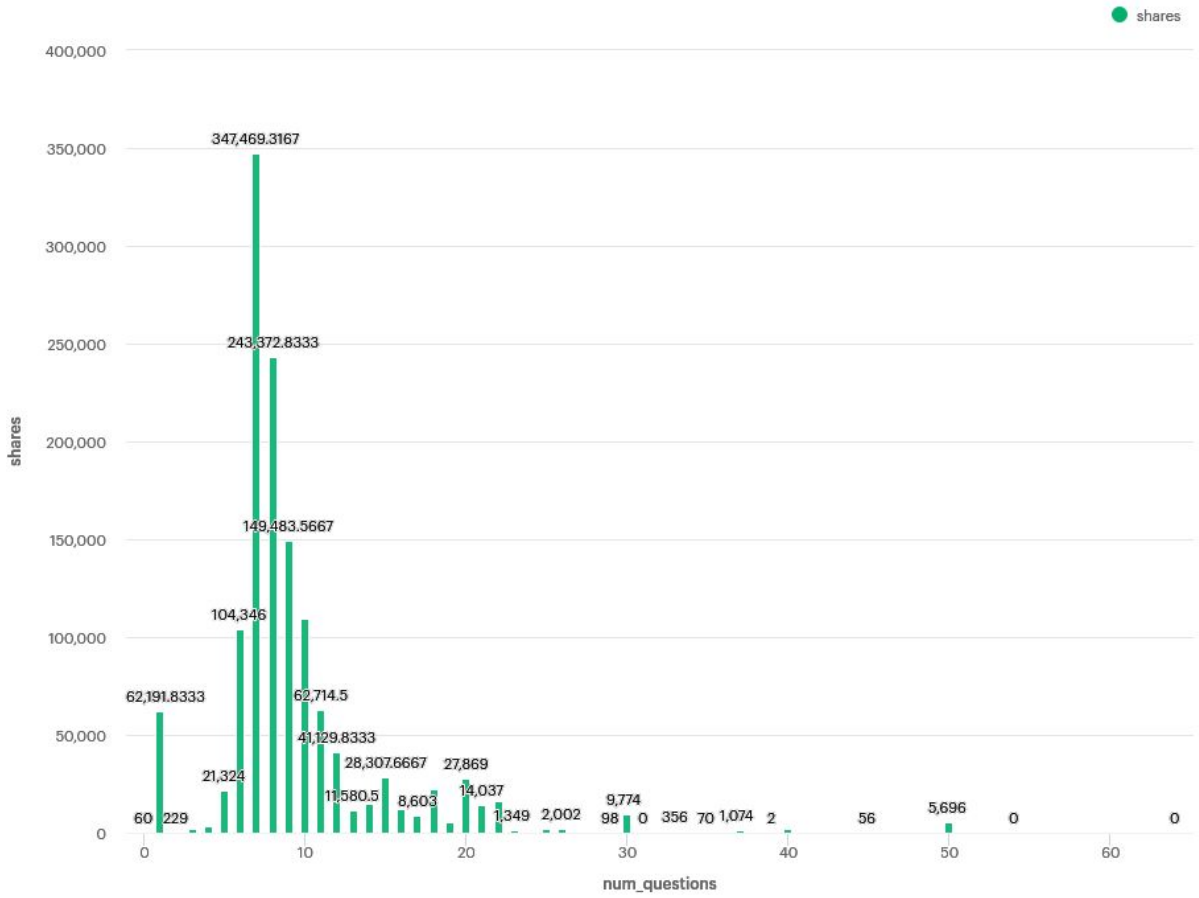
What is the optimal length for a quiz?

```
select num_questions, num_completes/count(num_questions) AS completions,
num_shares/count(num_questions) as shares, num_views/count(num_questions) as
views
from quiz_metadata a
join daily_agg_quiz_metrics b
on a.quiz_id = b.quiz_id
GROUP BY num_questions, num_completes, num_shares, num_views
having (avg(num_completes) /  avg(num_views) ) < 1
order by completions desc
```

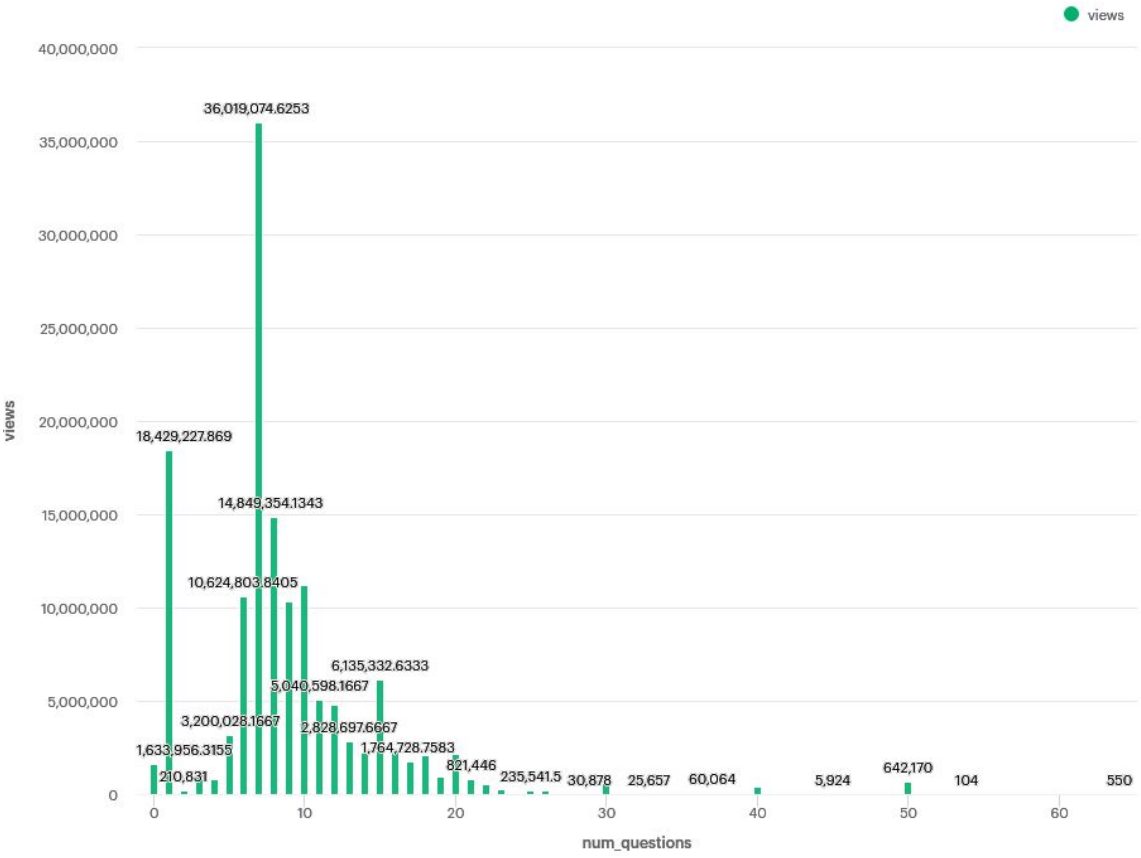
Completion rate



Share rate



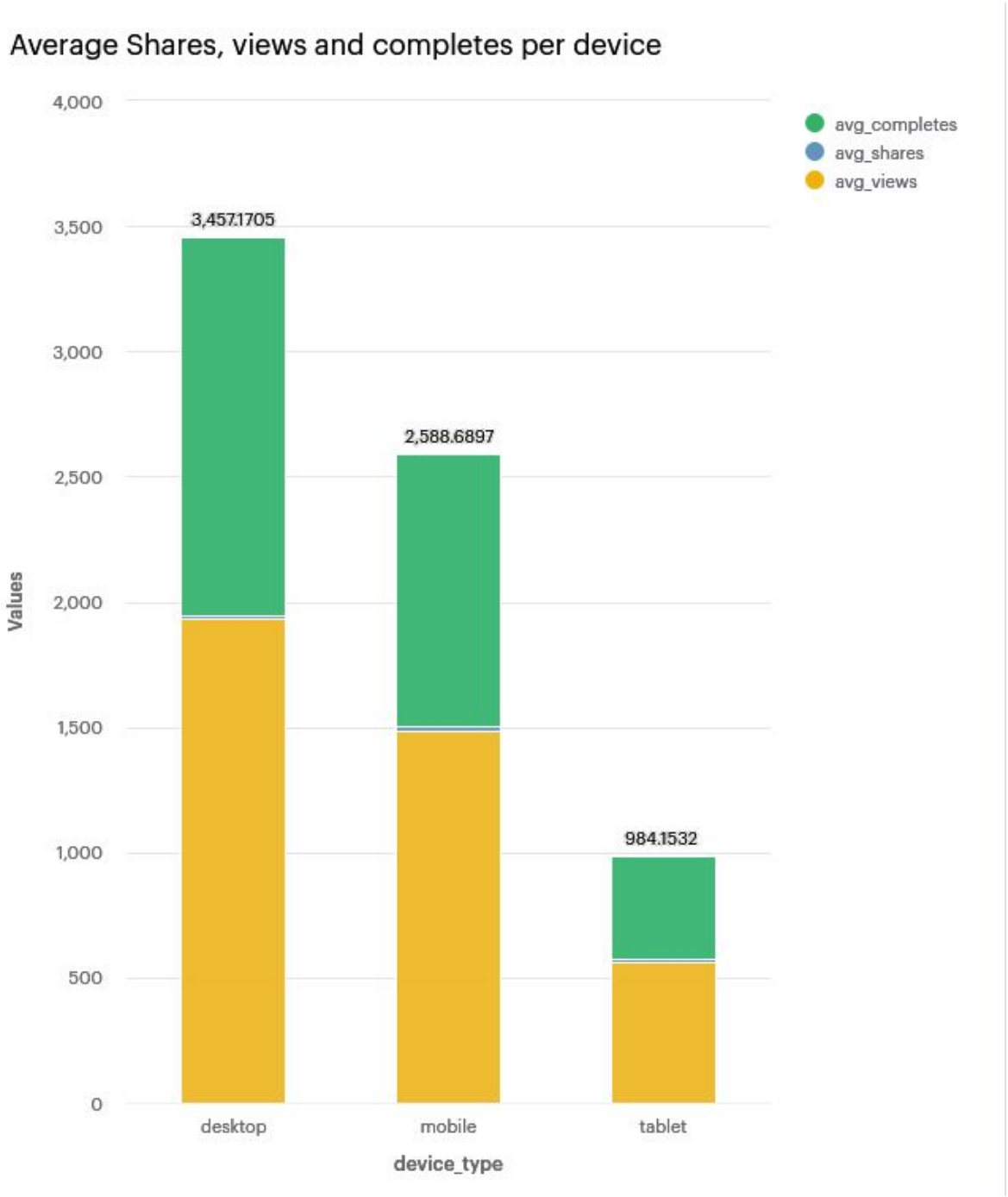
Views



From the graph above we can see that completion/share and views start to drop when the quiz exceeds 10 questions.
Optimal length of a quiz looks to be around 7-10 questions in length, which results in the highest number of completes and views and shares.

What are the differences between the device type? Should we focus on one more than the other?

```
Select count(device_type) as num_of_devices, device_type, avg(num_views) as
avg_views, avg(num_completes) as avg_completes, avg(num_shares) as avg_shares
from daily_agg_quiz_metrics a
join quiz_metadata b
on a.quiz_id = b.quiz_id
GROUP by device_type
```



The graph above shows desktop having the most views but mobile having the most shares. Completions are neck and neck for both desktop and mobile devices. Tablet is performing poorly in all three categories. I would suggest focusing on mobile and desktop.