Analytic case

Conditions

data: analytics_test_case

You will have the following data on user groups for Spring 2024:

- **date** - the date;
- **users** - the number of users in the group;
- **revenue** - the amount of revenue;
- **split_group** - the split group number. 0 - control group without changes, 1 - test group;
- **gender** - the gender of users;
- **os_version** - the version of the operating system;
- **age_group** - the age group of users;
- **country_group** - the user's country. Countries are grouped into several categories.

You are an analyst in a product IT company. There was a technical failure on the platform, which may have an impact on revenue. The failure occurred on May 16, 2024, and its consequences are ongoing. You need to assess the potential reduction in revenue due to the failure. At the same time, it is worth considering that many other factors can affect revenue.

Among them are:

- the structure of users and advertising budgets changed during this period;
- a major split test has been conducted since March 15;
- the support team was reduced at the end of March;
- the logic of working with emails changed at the beginning of April;
- a new OS release occurred in mid-April;
- revenue may exhibit cyclicality.

It is necessary to identify as many relevant factors as possible, evaluate and account for their impact. After that, determine how much the technical failure on May 16 specifically reduced revenue.

The criteria

The criteria by which the work will be assessed:

You can earn up to 20 points for:

- Identifying the trend (1) and evaluating its impact (1)
- Identifying monthly cyclicality (1) and evaluating its impact (1)
- Identifying weekly cyclicality (1) and evaluating its impact (1)
- Evaluating the impact of the split test (3)
- Evaluating the effect of the new OS launch (2)
- Identifying the change in the male/female structure (1) and evaluating its impact (1)
- Identifying a significant drop in one of the groups on May 20 (1) and evaluating its impact (1)
- Evaluating the impact of the technical failure on May 16 (1-3, depending on accuracy)
- The quality of the materials presented (2 points)

Query for aggregating daily revenue

For this case, we can use Descriptive Analytics, which explains what happened in the past. It allows us to verify existing facts, trends, and connections in the data, enabling us to answer questions about what happened, how often it happened, and what factors contributed to it happening.

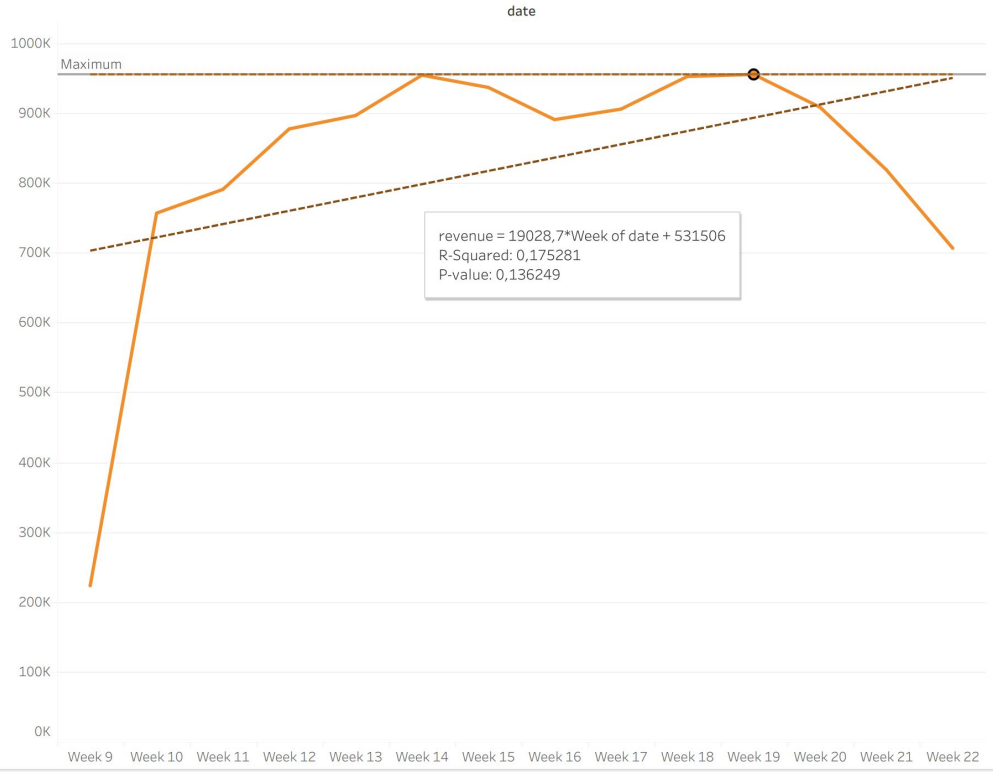
The trend was 0,26 %.

SQL:

```
SELECT date, SUM(revenue) AS daily_revenue
FROM analytics_test_case
GROUP BY date
ORDER BY date;
```

Query for aggregating weekly revenue

The results show the overall revenue trend by week from March to May 2024. The highest revenue was observed in the 19th week (early May), followed by a decline.



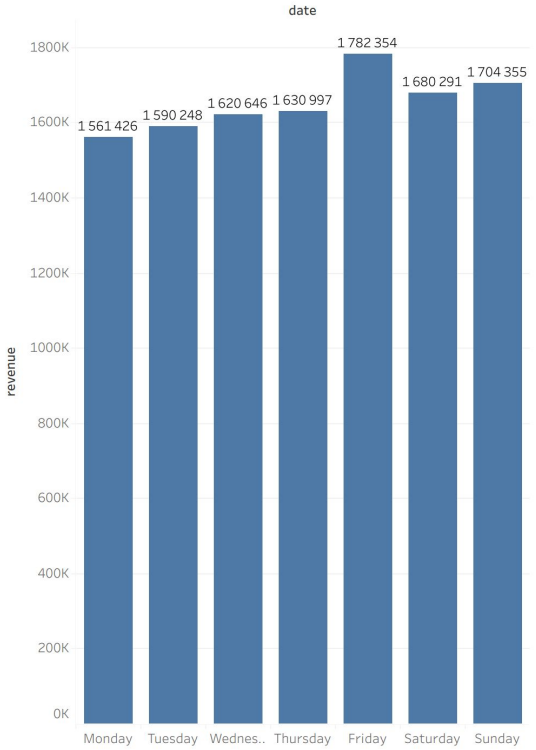
SQL:

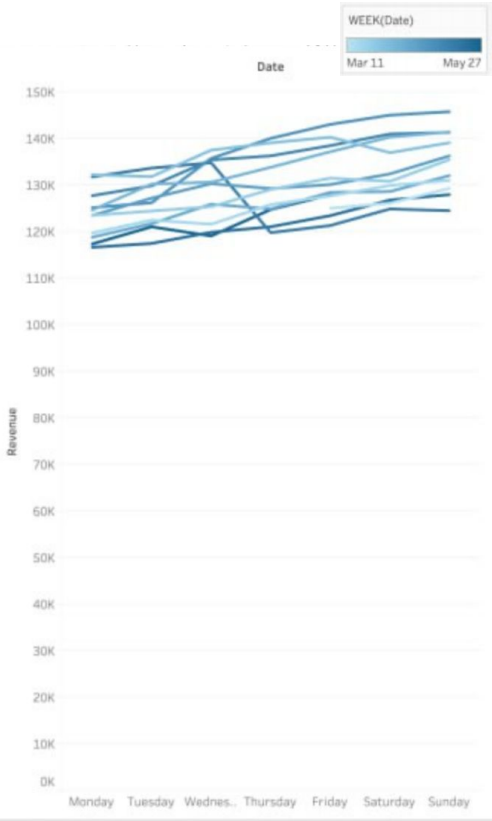
```
SELECT
 TO_CHAR(DATE_TRUNC('week', date::date),
 'MM-DD-YYYY') AS week_start,
 SUM(revenue) AS weekly_revenue
FROM analytics_test_case
GROUP BY week_start
ORDER BY week_start;
```

Query for aggregating revenue by days of the week

The data shows the distribution of revenue by days of the week. There is a slight cyclicality observed, with peaks occurring before the weekends (on Friday).

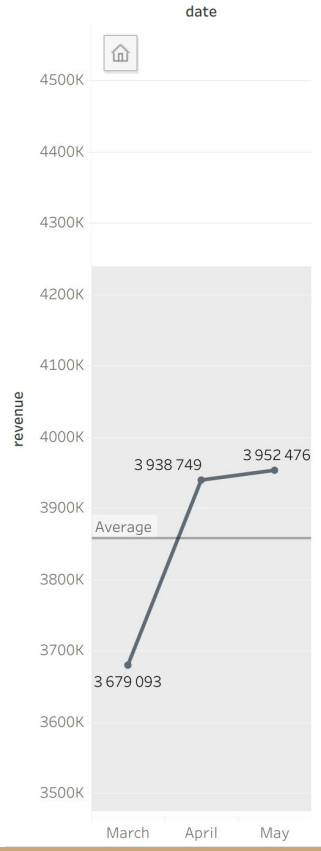
Weekly cyclicality is observed: less revenue on Mondays and more towards the end of the week. The week is "broken" because a few days before the start of the test fall into it.





Query for aggregating monthly revenue

Revenue gradually increases from March to May. The highest revenue is observed in May. Average is 3 856 773.



SQL:

```
SELECT
 TO_CHAR(DATE_TRUNC('month', date::date), 'DD-MM-YYYY') AS month_start,
 SUM(revenue) AS monthly_revenue
FROM analytics_test_case
GROUP BY month_start
ORDER BY month_start;
```

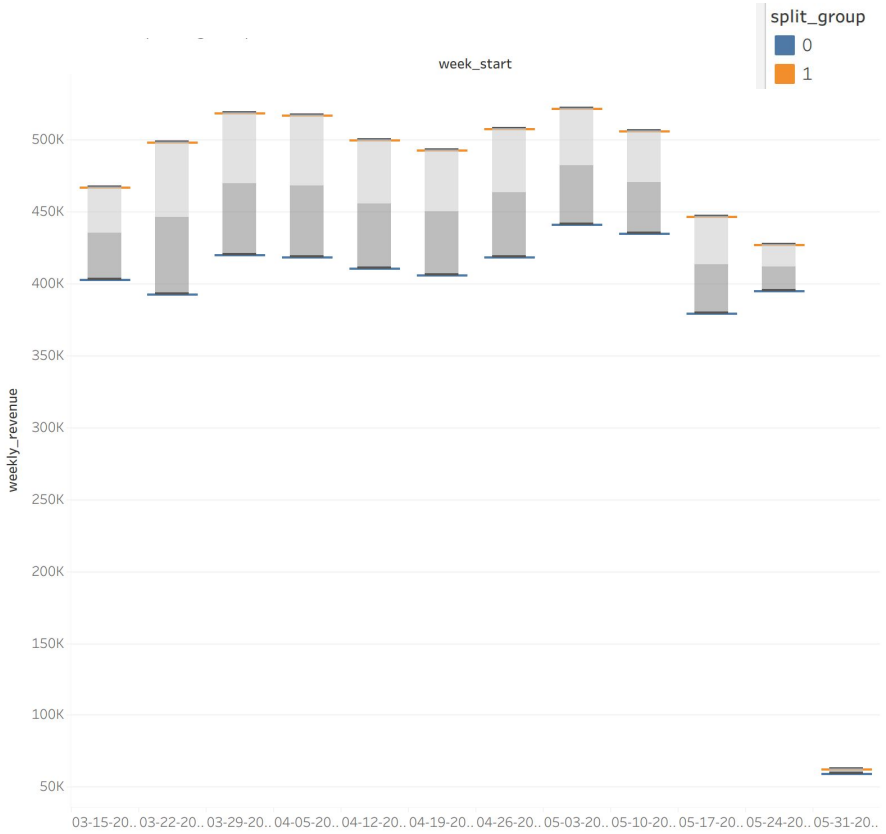
Evaluation of the split impact with revenue aggregation by weeks

- Average revenue for group 0: 381859.10
- Average revenue for group 1: 455772.48

T-test for assessing the statistical significance of the difference

- t-statistic: -1.5633
- p-value: 0.1323

The p-value is greater than 0.05, so we cannot reject the null hypothesis. This means there is not enough evidence to claim that the average weekly revenue between the two groups differs in a statistically significant way. This may indicate that the changes implemented in the test group did not have a significant impact on revenue.

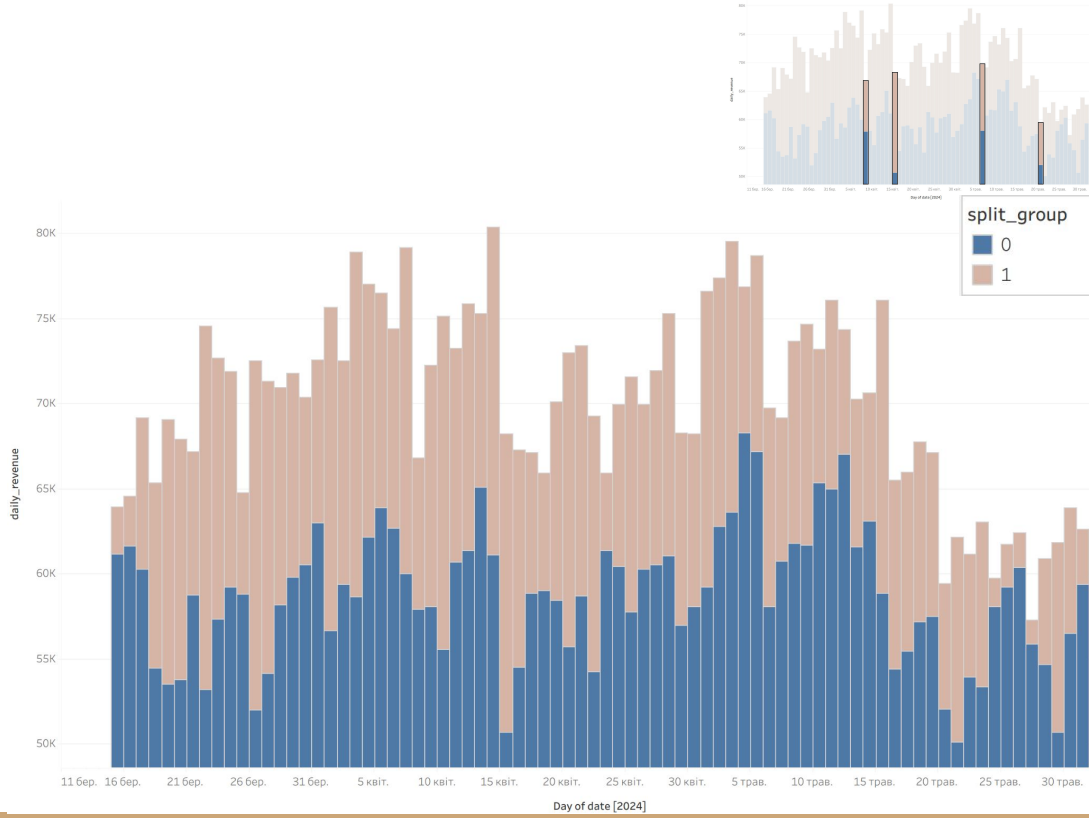


SQL:

```
SELECT
 split_group,
 TO_CHAR(DATE_TRUNC('week', date::timestamp - INTERVAL '4 days') +
INTERVAL '4 days', 'MM-DD-YYYY') AS week_start,
 SUM(revenue) AS weekly_revenue
FROM analytics_test_case
WHERE date >= '2024-03-15'
GROUP BY split_group, week_start
ORDER BY week_start;
```

Evaluation of the split impact with revenue aggregation by days

Significant revenue drops were found in both split groups. On April 15, there was a drop in both split groups (0 - by 10456, 1 - by 12173), likely related to the launch of the new OS. On April 8 (0 - only by 2116, but 1 - by 9111), likely due to changes in the email logic. Additionally, on May 6 (0 - by 9111, 1 - by 8943) and May 20 (0 - by 5411, 1 - by 7708).



SQL:

```
SELECT
 split_group,
 date,
 SUM(revenue) AS daily revenue
FROM analytics test case
WHERE date > '2024-03-15'
GROUP BY split_group, date
ORDER BY date;
```

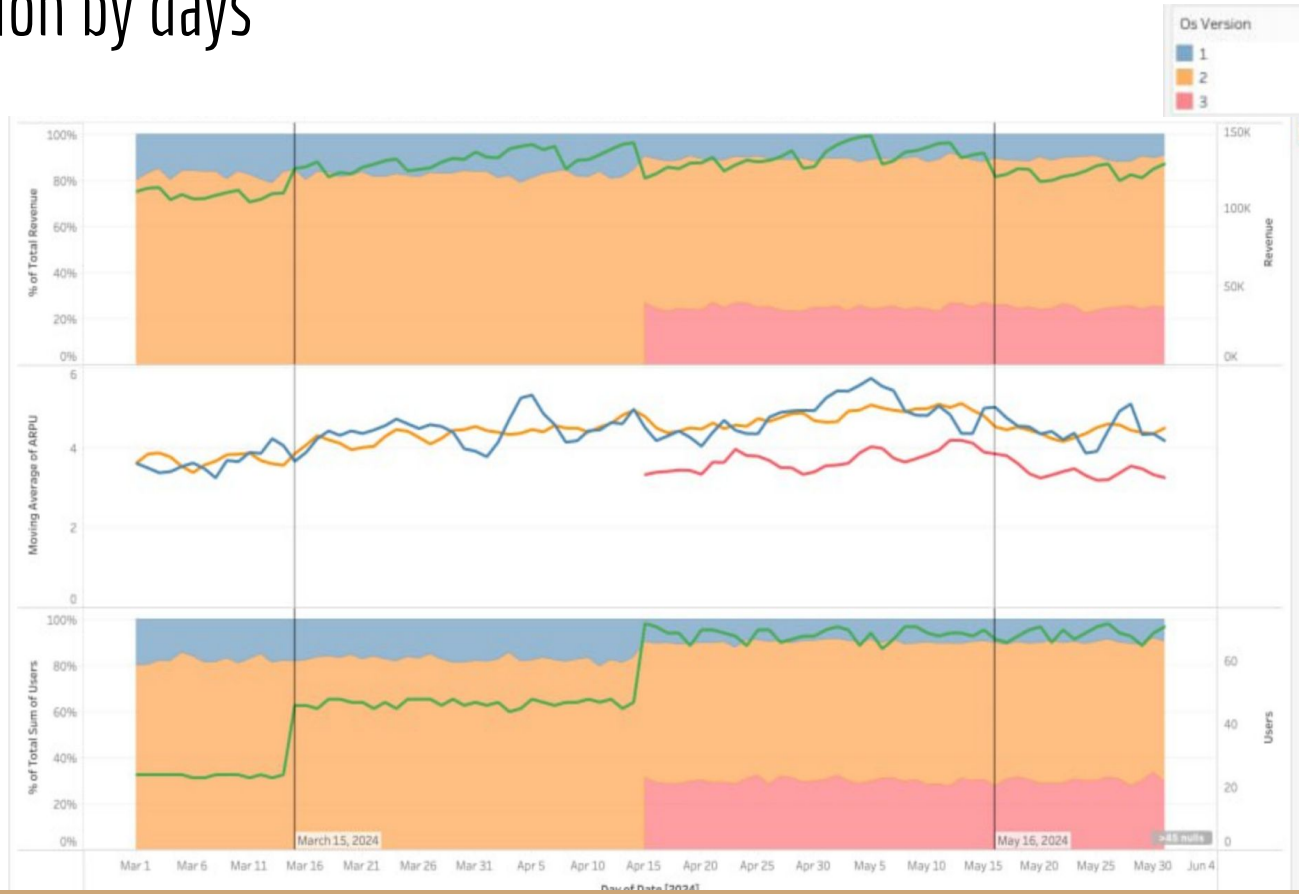
Evaluation of the effect of the new OS launch

with revenue aggregation by days

The visualizations collectively show how revenue, ARPU, and user distribution have changed over time for each OS version, with a specific focus on the impact of introducing OS Version 3.

On April 15, a third OS with worse ROAS was introduced, which caused a revenue drop. This was later compensated by the increase in ROAS of the first OS.

We can conclude that we are losing money, because of pouring ARPU after adding the third OS.



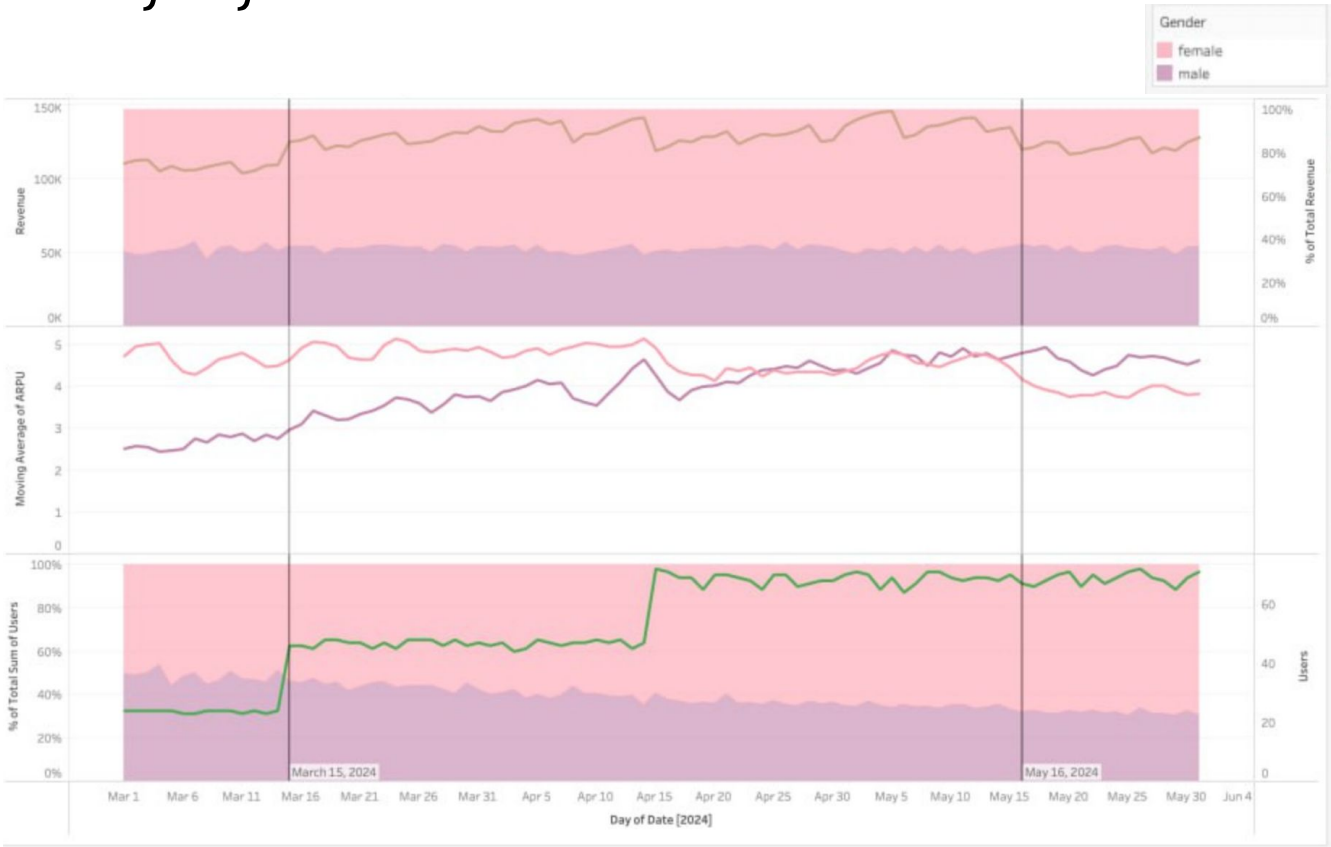
SQL:

```
SELECT
 os_version,
 date,
SUM(revenue) AS daily_revenue
FROM analytics_test_case
GROUP BY os_version, date
ORDER BY  date;
```

Identification of the change in the male/female structure

with revenue aggregation by days

On May 15, the ARPU (Average Revenue Per User) for women decreased (revenue from them did not drop because the number of women increased), but overall revenue fell because the number of men with good ROAS (Return on Advertising Spend) decreased. Since March 15, the proportion of men has been decreasing, but their ROAS has been increasing.



SQL:

```
SELECT
 gender,
 date,
SUM(revenue) AS daily revenue
FROM analytics_test_case
GROUP BY gender, date
ORDER BY  date;
```

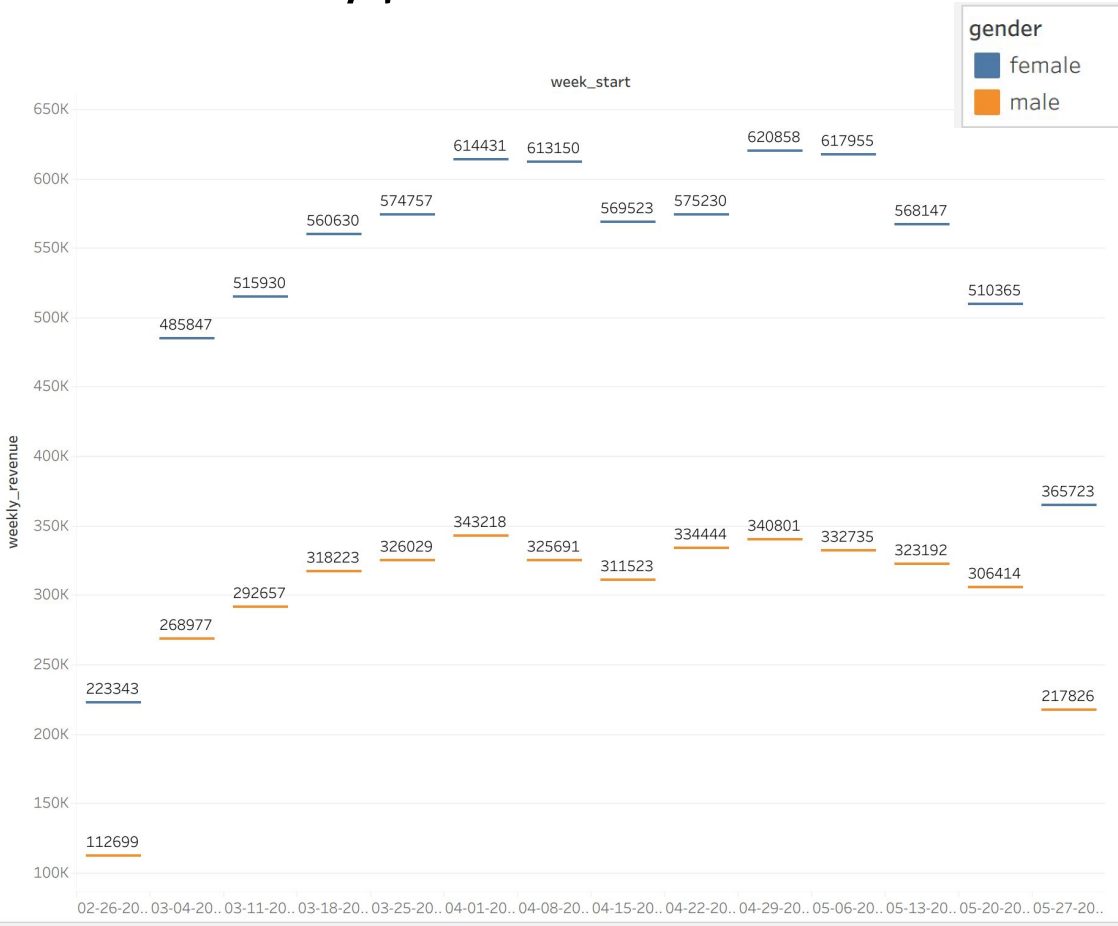
Identification of the change in the male/female structure with revenue aggregation by weeks

Average income:
Women: 529,706.27
Men: 296,745.01

T-test results:
T-statistic: -6.82
P-value: 3.09 × 10^(-7)

Interpretation:

The average weekly income for women is significantly higher than for men. The result of the t-test shows a very low p-value, indicating that the difference in average income between genders is statistically significant.



SQL:

```
SELECT
 gender,
 TO_CHAR(DATE_TRUNC('week', date::timestamp),
'MM-DD-YYYY') AS week_start,
 SUM(revenue) AS weekly_revenue
FROM analytics_test_case
GROUP BY gender, week_start
ORDER BY week_start;
```

Evaluation of the impact of the technical failure on May 16

Conclusions:

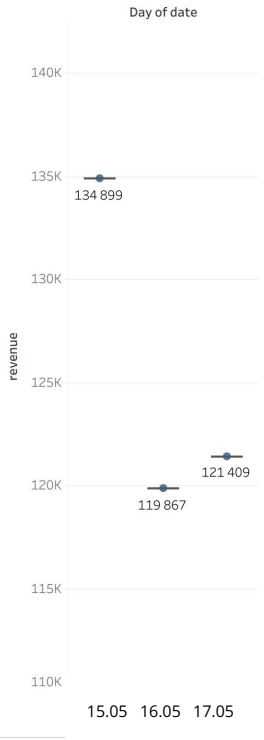
Decrease in revenue on May 16:
Revenue on May 16 was 119,866.87, which is 15,032.20 (approximately 11.1%) less than on May 15.

Slight recovery in revenue on May 17:
On May 17, revenue increased to 121,408.56, which is 1,541.69 (approximately 1.3%) more than on May 16, but still 13,490.51 (approximately 10%) less than on May 15.

Assessment of the impact of the technical failure:

Reduction in revenue on the day of the technical failure:
The technical failure on May 16 led to a significant revenue decrease of 11.1% compared to the previous day.

Prolonged impact:
On May 17, revenue partially recovered but remained 10% below the level of May 15. This may indicate that the technical failure had a prolonged negative impact on revenue.



SQL:

```sql
SELECT
 date,
 SUM(revenue) AS daily_revenue
FROM analytics_test_case
WHERE date BETWEEN '2024-05-15' AND '2024-05-17'
GROUP BY date
ORDER BY date;
```