## JDS Final Task - Dilan's Travel Guide - PART 2: Segmentation

# 1. First readers

1.1 First I check the daily number of first reads and I put it on a line chart to see the trend. **The number of daily new readers is growing**, which is a good sign for Dilan's business. There is no data for 2018-03-31.

```
SELECT DATE_TRUNC('day', my_time) AS day,
       COUNT(*) AS num_lines
FROM first_read
GROUP BY day
ORDER BY day;
```

1.2 **Dilan's blog has 210 020 first readers.**

```
SELECT COUNT(DISTINCT(user_id))
FROM first_read;
```

1.3 I check the most important source of new readers, it is Reddit. **Although it is not the most expensive source Dilan pays for, it pays well. Maybe he should spend less on AdWords and more on Reddit.**

```
SELECT source, COUNT(*) as num_first_read
FROM first_read
GROUP BY source
ORDER By num_first_read DESC;
```

1.4 I calculate the number of first readers by country. **Country_7 and country_2 outperform the others, so they can be good targets for more aggressive marketing.**

```
SELECT country, COUNT(*) as num_first_read
FROM first_read
GROUP BY country
ORDER By num_first_read DESC;
```

1.5 I count the popularity of the topics, Asia is the winner by far with almost twice as much reads that the second Europe. **It suggests that Dilan should put greater emphasis on topics related to Asia.**

```
SELECT topic, COUNT(*) as num_first_read
FROM first_read
GROUP BY topic
ORDER By num_first_read DESC;
```

1.6 I checked the relation between the source of readers and the topic they read. The pattern seems quite clear: AdWords: Europe, Reddit: Asia and SEO: North America. **It confirms that if he wants to focus on Asia, he should use Reddit more extensively.** But also, he knows which source he should turn to increase the number of users from continents respectively.

```
SELECT source, topic, COUNT(*) as num_first_read
FROM first_read
GROUP BY source, topic
ORDER By source, num_first_read DESC;
```

1.7 The country-topic segment shows similar pattern among countries: the main topics are Asia, Europe and North-America in every country. Related to topics there is no significant difference in preferences among countries.

```
SELECT country, topic, COUNT(*) as num_first_read
FROM first_read
GROUP BY country, topic
ORDER By country, num_first_read DESC;
```

1.8 In all country Reddit is the main source, so that related to sources there is no significant difference in preferences among countries.

```
SELECT country, source, COUNT(*) as num_first_read
FROM first_read
GROUP BY country, source
ORDER By country, num_first_read DESC;
```

1.9 There are no changes in the trend that the first readers come from Reddit, AdWords and the SEO. The situation is the same in the case of topics and countries.

```
SELECT DATE_TRUNC('day', my_time) AS day,
        source,
       COUNT(*) AS num_lines
FROM first_read
GROUP BY day, source
ORDER BY day, num_lines DESC;
```

## 2. Returning readers

2.1 I count the trend of the number of returning reads. It is fluctuating. **It has a peak on 2018-03-27 then suddenly drops.** (Contrary to the case of first readers we have some data for 31th March.)

```
SELECT DATE_TRUNC('day', my_time) AS day,
       COUNT(*) AS num_ret_read
FROM ret_read
GROUP BY day
ORDER BY day;
```

2.2 I check the number of readers (by user_id) and I see a similar trend in the number of reads. Comparing these two indicators I can see the number of pages read by on a given day. Mostly it varies between 1 and 2.

```
SELECT DATE_TRUNC('day', my_time) AS day,
    COUNT(DISTINCT(user_id)) AS num_userid
FROM ret_read
GROUP BY day
ORDER BY day;
```

2.3 I want to know readers from which country are the most active: country_5. **It seems that new readers rather come from country_7 and country_2 (country_5 is only the 3rd on this list), but readers from country_5 are the most loyal ones. So now it seems they are worth focusing on.** (Country_7 and country_2 are 2nd and 3rd on this list so that they are still important.)

```
SELECT country, COUNT(*) as num_ret_read
FROM ret_read
GROUP BY country
ORDER By num_ret_read DESC;
```

2.4 Reddit still outperforms the others as source of returning readers but SEO is not far behind.

```
SELECT source, COUNT(*) as num_ret_read
FROM first_read
JOIN ret_read
ON first_read.user_id = ret_read.user_id
GROUP BY source
ORDER BY num_ret_read DESC;
```

2.4 **The most liked topic is still Asia** but in this case North_America definitely outperforms Europe in the 2nd place on the list. The distribution of reads changed too, **Asia's dominance decreased a lot.**

```
SELECT topic, COUNT(*) as num_ret_read
FROM ret_read
GROUP BY topic
ORDER By num_ret_read DESC;
```

2.5 In case of country_1 and country_8 Noth_America is the most popular topic, in the others Asia is on top.

```
SELECT country, topic, COUNT(*) as num_ret_read
FROM ret_read
GROUP BY country, topic
ORDER By country, num_ret_read DESC;
```

2.6 It is interesting to see that before 2018-01-27 the topic North_America was the most popular amond returning readers, then it has turned to Asia.

```
SELECT DATE_TRUNC('day', my_time) AS day,
       topic,
       COUNT(*) AS num_ret_read
FROM ret_read
GROUP BY day, topic
ORDER BY day, num_ret_read DESC;
```

2.7 At the start of the blog Dilan had the most returning readers from country_4 but after some days it definitely changed to country_5.

```
SELECT DATE_TRUNC('day', my_time) AS day,
       country,
       COUNT(*) AS num_ret_read
FROM ret_read
GROUP BY day, country
ORDER BY day, num_ret_read DESC;
```

2.8 **Dilan's blog has 66 231 returning readers. Less than one-third of the first readers becomes returning readers which ratio needs to improve.**

```
SELECT COUNT(DISTINCT(user_id))
FROM ret_read;
```

2.9 **It is good news that the returning readers' less than 20% return only for 1 another read.** There are returning readers above 50 reads! **The average is number of read of returning readers is 5,6.** It seems that the readers like returning to Dilan's blog, with a good content management the number of returns could be even bigger.

```
SELECT
    avg_reads_per_user,
    COUNT(user_id) AS num_users
FROM
    (SELECT user_id,
        AVG(num_reads) AS avg_reads_per_user
    FROM (SELECT user_id,
            COUNT(*) AS num_reads
        FROM ret_read
```

```
        GROUP BY user_id) AS reads_user
    GROUP BY user_id) AS avg_reads_user
GROUP BY avg_reads_per_user
ORDER BY avg_reads_per_user;
```

## 3. Subscribers

3.1 **The number of subscribers 7618, which is 3.6% of the first readers and 12% of the returning readers. It means that only one-sixth of even the returning readers subscribe. It definitely needs to be increased.**

```
SELECT COUNT(*)
FROM subs;
```

3.2 **The trend of the number of new subscribers is fluctuating but not significantly growing, which is not good for the business. And again, the is no data on 31th March, the reason needs to be detected.**

```
SELECT DATE_TRUNC('day', my_time) AS day,
       COUNT(*) AS num_lines
FROM ret_read
GROUP BY day
ORDER BY day;
```

3.3 **Readers from country_5 have the most subscribers.**

```
SELECT country,
COUNT(*) AS num_subs
FROM first_read
JOIN subs
ON first_read.user_id = subs.user_id
GROUP BY country
ORDER BY num_subs DESC;
```

3.4 **Reddit is the main source of subscribers but SEO is also very important.**

```
SELECT source,
COUNT(*)
FROM first_read
JOIN subs
ON first_read.user_id = subs.user_id
GROUP BY source
ORDER BY count DESC;
```

3.5 **By far Asia is the most interesting topic among subscribers.**

```
SELECT topic,
COUNT(*)
FROM first_read
JOIN subs
ON first_read.user_id = subs.user_id
GROUP BY topic
ORDER BY count DESC;
```

3.6 **97% of subscribers comes from returning readers so that to develop the business Dilan should increase the number of returning readers to get more subscriptions.**

```
SELECT COUNT(DISTINCT(user_id))
FROM
    (SELECT subs.user_id,
    COUNT(*)
    FROM ret_read
    JOIN subs
    ON ret_read.user_id = subs.user_id
    GROUP BY subs.user_id) AS num_reads_by_subs;
```

3.7 **The average number of reads among subscribers is 15.8 which shows that they are really fanatic about Dilan's travel blog.**

```
SELECT AVG(count)
FROM
    (SELECT subs.user_id, COUNT(*)
    FROM ret_read
    JOIN subs
    ON ret_read.user_id = subs.user_id
    GROUP BY subs.user_id) AS reads_by_subs;
```

## 4. Buyers

4.1 Dilan's blog has 6648 buyers and 8407 purchases. **Interesting to see that most of the subscribers (87%) become buyers.**

```
SELECT COUNT(*)
FROM buy;
```

```
SELECT COUNT(DISTINCT(user_id))
FROM buy;
```

4.2 The number of purchases is fluctuating but the overall trend is promising. Again, something happened after 27th March as the figure drops.

```
SELECT DATE_TRUNC('day', my_time) AS day,
       COUNT(*) AS num_lines
FROM buy
GROUP BY day
ORDER BY day;
```

4.3 Not surprisingly country_5 is the dominant among buyers too...

```
SELECT country,
COUNT(*)
FROM first_read
JOIN buy
ON first_read.user_id = buy.user_id
GROUP BY country
ORDER BY count DESC;
```

4.4 ... and also Reddit is the main source...

```
SELECT source,
COUNT(*)
FROM first_read
JOIN buy
ON first_read.user_id = buy.user_id
GROUP BY source
ORDER BY count DESC;
```

4.5 ...and Asia is the most attractive topic.

```
SELECT topic,
COUNT(*)
FROM first_read
JOIN buy
ON first_read.user_id = buy.user_id
GROUP BY topic
ORDER BY count DESC;
```

## 5. Bugdet

5.1 There are 4881 buyers who buy only the $8 e-book, it means $39 408 as income.

```
SELECT COUNT(goods8.user_id) AS num_users_only_8
FROM
    -- Users who bought goods priced $8
    (SELECT DISTINCT user_id
    FROM buy
    WHERE price = '8') AS goods8
LEFT JOIN
    -- Users who bought goods priced $80
    (SELECT DISTINCT user_id
    FROM buy
    WHERE price = '80') AS goods80
ON goods8.user_id = goods80.user_id
WHERE goods80.user_id IS NULL;
```

5.2 There are only 8 buyere who buy only the $80 video course, it equals $640 as income.

```
SELECT COUNT(goods80.user_id) AS num_users_only_80
FROM
    -- Users who bought goods priced $80
    (SELECT DISTINCT user_id
    FROM buy
    WHERE price = '80') AS goods80
LEFT JOIN
    -- Users who bought goods priced $8
    (SELECT DISTINCT user_id
    FROM buy
    WHERE price = '8') AS goods8
ON goods80.user_id = goods8.user_id
WHERE goods8.user_id IS NULL;
```

5.3 There are 1759 users who buy both products, it gives $154 792 income for Dilan.

```
SELECT COUNT(goods8.user_id) AS num_users_both
FROM
    -- Users who bought goods priced $8
    (SELECT DISTINCT user_id
    FROM buy
    WHERE price = '8') AS goods8
INNER JOIN
    -- Users who bought goods priced $80
    (SELECT DISTINCT user_id
    FROM buy
    WHERE price = '80') AS goods80
ON goods8.user_id = goods80.user_id;
```

5.4 In 95% of cases when users buy both products they buy the $8 e-book first then they decide to buy the $80 video course.

```
SELECT COUNT(*)
FROM
    (SELECT goods8.my_time, goods8.price, goods80.price, goods80.my_time
        FROM
            --number of buyers with goods priced 8 dollars
            (SELECT *
            FROM buy
            WHERE price = '8') AS goods8
        FULL JOIN
            --number of buyers with goods priced 80 dollars
            (SELECT *
            FROM buy
            WHERE price = '80') AS goods80
        ON goods8.user_id = goods80.user_id
WHERE goods8.price = '8' AND goods80.price = '80' AND goods8.my_time <
goods80.my_time) as full_table;
```

5.5 Dilan's overall income from the blog is $194 480.

```
SELECT SUM(price)
FROM buy;
```

5.6 The trend of daily income fluctuates, the last week of March seems successful but the figure drops after 2018-03-27. There is no data on 2018-03-31.

```
SELECT DATE_TRUNC('day', my_time) AS day,
    SUM(price) AS sum_prices
FROM buy
GROUP BY day
ORDER BY day;
```