User Analysis

1 – Identifying Repeat Visitors

Analyze repeat visits to understand overall user behavior and identify high value customers.

Assignment: Pull data on how many of website visitors come back for another sessions.

```sql
-- Identify Repeat Visitors
WITH t1 AS -- Step 1: Pull new sessions associated with each user
(
    SELECT user_id, website_session_id
    FROM website_sessions
    WHERE created_at < '2014-11-01' -- the date of the assignment
    AND created_at >= '2014-01-01' -- prescribed date range in assignment (2014 to date)
    AND is_repeat_session = 0 -- new sessions only
) ,

t2 AS -- Step 2: Use the user_id from Step 1 to find any repeat sessions those users had
(SELECT t1.user_id,
        t1.website_session_id AS new_session_id,
        ws.website_session_id AS repeat_session_id
FROM t1 LEFT JOIN website_sessions ws
  ON t1.user_id = ws.user_id
  AND ws.is_repeat_session = 1 -- was a repeat session
  AND ws.website_session_id > t1.website_session_id -- session occurred later than new session
  AND ws.created_at < '2014-11-01' -- the date of the assignment
  AND ws.created_at >= '2014-01-01' -- prescribed date range in assignment
),

t3 AS -- Step 3: Analyze data at User Level (#sessions for each user)
(SELECT user_id,
        COUNT(DISTINCT new_session_id) AS new_sessions,
        COUNT(DISTINCT repeat_session_id) AS repeat_sessions
FROM t2
GROUP BY 1 ORDER BY 3 DESC) -- user_level table (Step 3 result)

-- Step 4: Aggregate the user_level analysis to generate overall Behavior Analysis
SELECT repeat_sessions, COUNT(DISTINCT user_id) AS users
FROM t3 GROUP BY 1;
```

repeat_sessions	users
0	126804
1	14084
2	316
3	4685

- As shown in the breakdown, 126,804 users had no repeat sessions; 14,084 users had only one repeat sessions, and 4,685 users had three. It's worth noting that very few users (316) had two repeat sessions.
- Since quite a few customers are coming back for a repeat visit, it'd be useful to dig deeper into repeat customer behavior and understand customer value.

2 – Analyzing Time to Repeat (Repeat Behavior)

Assignment: Understand the behavior of these repeat customers by analyzing the minimum, maximum, and average time between the first and second session for customer who do come back.

```sql
-- Analyzing Time to repeat
WITH t1 AS -- Step 1: Identify the relevant new sessions (new sessins table)
(
  SELECT user_id,
      website_session_id,
      created_at
  FROM website_sessions
  WHERE created_at < '2014-11-03' -- the date of the assignment
    AND created_at >= '2014-01-01' -- prescribed date range in assignment
    AND is_repeat_session = 0 -- new sessions only
), -- new session table (step 1 result)

t2 AS  -- Step 2: use user_id from Step 1 to find any repeat sessions those users had
(
SELECT t1.user_id,
      t1.website_session_id AS new_session_id,
      t1.created_at AS new_session_created_at,
      s.website_session_id AS repeat_session_id,
      s.created_at AS repeat_session_created_at
FROM t1 LEFT JOIN website_sessions s
  ON t1.user_id = s.user_id
  AND s.is_repeat_session = 1 -- repeat session
  AND s.website_session_id > t1.website_session_id -- session occurred later than new session
  AND s.created_at < '2014-11-03' -- the date of the assignment
  AND s.created_at >= '2014-01-01' -- prescribed date range in assignment
), -- sessions_w_repeats_for_time_diff table

t3 AS -- Step 3: Find the created_at times for 1st and 2nd sessions
(
SELECT user_id, new_session_id, new_session_created_at,
      MIN(repeat_session_id) AS second_session_id,
      MIN(repeat_session_created_at) AS second_session_created_at
FROM t2
WHERE repeat_session_id IS NOT NULL
GROUP BY 1,2,3),

t4 AS -- Step 4: Find the difference between 1st and 2nd sessions at a User level
(
SELECT user_id,
      DATEDIFF(second_session_created_at, new_session_created_at) AS days_first_to_second_session
FROM t3) -- user_first_to_second (user level table)

-- Step 5: Aggregate the user level data to find the AVG, MIN, MAX
SELECT AVG(days_first_to_second_session) AS avg_days_first_to_second,
      MIN(days_first_to_second_session) AS min_days_first_to_second,
      MAX(days_first_to_second_session) AS max_days_first_to_second
FROM t4; -- aggregate the user level data
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

avg_days_first_to_second	min_days_first_to_second	max_days_first_to_second
33.2651	1	69

3 – Analyzing Repeat Channel

Assignment: Pull data to show the channels that repeat customer come back through on their subsequent sessions. Compare new vs. repeat sessions by channel to understand whether repeat customers are coming back through a free channel or if you're paying again to get the customers through paid channels.

```
SELECT utm_source, utm_campaign,http_referer,
    COUNT(CASE WHEN is_repeat_session = 0 THEN website_session_id ELSE NULL END) AS new_sessions,
    COUNT(CASE WHEN is_repeat_session = 1 THEN website_session_id ELSE NULL END) AS repeat_sessions
FROM website_sessions
WHERE created_at < '2014-11-05' -- the date of the assignment
  AND created_at >= '2014-01-01' -- prescribed date range in assignment
GROUP BY 1,2,3
ORDER BY 5 DESC;
```

utm_source	utm_campaign	http_referer	new_sessions	repeat_sessions
NULL	NULL	NULL	6591	10564
NULL	NULL	https://www.gsearch.com	5738	9450
gsearch	brand	https://www.gsearch.com	5196	8897
bsearch	brand	https://www.bsearch.com	1236	2130
NULL	NULL	https://www.bsearch.com	1401	2057
gsearch	nonbrand	https://www.gsearch.com	100278	0
bsearch	nonbrand	https://www.bsearch.com	19672	0
socialbook	pilot	https://www.socialbook.com	5095	0
socialbook	desktop_targeted	https://www.socialbook.com	2557	0

- The channel that drives the most repeat sessions is Direct-Type-In (with all null parameters).
- The second most repeat sessions are coming in through Organic Gsearch, closely followed by is Gsearch Brand campaign. Another few are coming in through Bsearch (brand and organic).
- No repeat sessions are coming in through other paid channels (Social paid or nonbrand search campaigns), which is a good thing as *paid brand is cheaper than paid nonbrand*.

Show clearer results with Simplified Channel Grouping:

```
SELECT
    CASE WHEN utm_source IS NULL AND http_referer IN ('https://www.gsearch.com', 'https://www.bsearch.com')
    THEN 'organic_search'
        WHEN utm_campaign = 'nonbrand' THEN 'paid_nonbrand'
        WHEN utm_campaign = 'brand' THEN 'paid_brand'
        WHEN utm_source IS NULL AND http_referer IS NULL THEN 'direct_type_in'
        WHEN utm_source = 'socialbook' THEN 'paid_social'
    END AS channel_group,
    COUNT(CASE WHEN is_repeat_session = 0 THEN website_session_id ELSE NULL END) AS new_sessions,
    COUNT(CASE WHEN is_repeat_session = 1 THEN website_session_id ELSE NULL END) AS repeat_sessions
FROM website_sessions
WHERE created_at < '2014-11-05' -- the date of the assignment
  AND created_at >= '2014-01-01' -- prescribed date range in assignment
GROUP BY 1
ORDER BY repeat_sessions DESC;
```

channel_group	new_sessions	repeat_sessions
organic_search	7139	11507
paid_brand	6432	11027
direct_type_in	6591	10564
paid_nonbrand	119950	0
paid_social	7652	0

- All in all, the company is not paying too much for extra visits.

4 – Analyzing New & Repeat Conversion Rates

Assignment: Use a comparison of New vs. Repeat session to order conversion rates and revenue per session to understand repeat performance.

```sql
-- Analyzing New & Repeat Conversion Rates
SELECT
    is_repeat_session,
    COUNT(DISTINCT s.website_session_id) AS sessions,
    COUNT(DISTINCT o.order_id)/COUNT(DISTINCT s.website_session_id) AS conv_rate,
    SUM(price_usd)/COUNT(DISTINCT s.website_session_id) AS rev_per_session
FROM website_sessions s LEFT JOIN orders o
ON s.website_session_id = o.website_session_id
WHERE s.created_at < '2014-11-08' -- date of the assignment
  AND s.created_at >= '2014-01-01' -- prescribed date range in assignment
GROUP BY 1;
```

Result Grid | Filter Rows: | Export:

is_repeat_session	sessions	conv_rate	rev_per_session
0	149787	0.0680	4.343754
1	33577	0.0811	5.168828

- Repeat sessions convert better to orders than new sessions.
- Repeat sessions also generate more revenue per sessions than new sessions.
- Customers who are familiar with your business or products tend to purchase a bit more.