1. Data Ingestion

To do import the data, I used the SSIS (Visual Studio) tool to ingest data from the flat file into the SQL Server.

|  |
| --- |
|  |

|  |  |
| --- | --- |
| № | SQL Queries for the questions |
| 1 | -- 1. How many action records are present for each type major?  SELECT  U1.action\_type\_major,  COUNT (DISTINCT action\_mysql\_id) NumberOfAction\_ForEachMajor  FROM [dbo].[User\_Activity] AS U1  GROUP BY U1.action\_type\_major  ; |
| 2 | -- 2. What is the time range of the data set?  SELECT MIN (session\_start\_date\_time) MIN\_SESSION\_DATE, MAX(session\_end\_date\_time) MAX\_SESSION\_DATE  FROM [dbo].[User\_Activity] WITH (NOLOCK)  WHERE session\_start\_date\_time != '1970-01-01 00:00:00.000' |
| 3 | -- 3. What is the average, minimum, and maximum number of action records on each day? A day should be  ---- considered midnight to midnight in EDT (Eastern Daylight Time, UTC-4).  SELECT  AVG(NUMBER\_OF\_ACTION\_RECORDS) [Avg of action records on each day],  MIN(NUMBER\_OF\_ACTION\_RECORDS) [Max of action records on each day],  MAX(NUMBER\_OF\_ACTION\_RECORDS) [Min of action records on each day]  FROM  (  SELECT  CAST (action\_request\_time AS DATE ) AS ACTION\_DAY,  COUNT(\*) NUMBER\_OF\_ACTION\_RECORDS  FROM [dbo].[User\_Activity] AS U1  GROUP BY CAST (action\_request\_time AS DATE )  ) AS T1  ; |
| 4 | -- 4. What is the average dwell time per page view?  -- Using aggregation analytic functions with CTE  WITH CTE\_1 AS (  SELECT  U1.\*,  AVG(ABS( CAST (DATEDIFF(MINUTE, U1.action\_request\_time, U1.action\_invisible\_time ) AS BIGINT))) OVER (PARTITION BY U1.action\_resource\_major ) AVG\_DWELL\_TIME,  ROW\_NUMBER () OVER (PARTITION BY U1.action\_resource\_major ORDER BY U1.action\_resource\_major) AS RN  FROM [dbo].[User\_Activity] AS U1  WHERE U1.action\_type\_major = 'V'  )  SELECT action\_resource\_major, AVG\_DWELL\_TIME FROM CTE\_1  WHERE RN = 1  -- using direct group by command  SELECT  U1.ACTION\_RESOURCE\_MAJOR,  ABS(SUM( CAST (DATEDIFF(MINUTE, U1.action\_request\_time, U1.action\_invisible\_time )AS BIGINT)) / COUNT(\*)) DWELL\_TIME\_BY\_MINUTE  FROM [dbo].[User\_Activity] AS U1  WHERE U1.action\_type\_major = 'V'  GROUP BY U1.ACTION\_RESOURCE\_MAJOR |
| 5 | -- 5. What is the average number of page views per session and per device?  SELECT AVG(T1.COUNT\_OF\_PAGE\_VIEW) [Avg number of page views for per device]  FROM  (  SELECT device\_id, COUNT (\*) COUNT\_OF\_PAGE\_VIEW  FROM [dbo].[User\_Activity] AS U1  WHERE U1.action\_type\_major = 'V'  GROUP BY device\_id  )AS T1  SELECT AVG(T1.COUNT\_OF\_PAGE\_VIEW) [Avg number of page views for per session]  FROM  (  SELECT session\_id, COUNT (\*) COUNT\_OF\_PAGE\_VIEW  FROM [dbo].[User\_Activity] AS U1  WHERE U1.action\_type\_major = 'V'  GROUP BY session\_id  )AS T1 |
| 6 |  |

1. Data Analysis

* Question 1: What times of the day is lowest and highest usage for each major type of actions?

SQL Query:

|  |
| --- |
| SELECT CAST (T1.[Hours] AS NVARCHAR (50)) + ':00',  SUM (PAGE\_VIEW) AS PAGE\_VIEW,  SUM (PUSH\_VIEW) AS PUSH\_VIEW,  SUM (PUSH\_CLICK) AS PUSH\_CLICK,  SUM (BANNER\_WAS\_VIEWED) AS BANNER\_WAS\_VIEWED,  SUM (BANNER\_WAS\_CLICKED) AS BANNER\_WAS\_CLICKED  FROM (  SELECT  DATEPART(HOUR, action\_request\_time) [Hours],  CASE WHEN ACTION\_TYPE\_MAJOR = 'AD\_BAN\_CLICK' THEN COUNT (\*) ELSE 0 END BANNER\_WAS\_CLICKED,  CASE WHEN ACTION\_TYPE\_MAJOR = 'AD\_BAN\_IMP' THEN COUNT (\*) ELSE 0 END BANNER\_WAS\_VIEWED,  CASE WHEN ACTION\_TYPE\_MAJOR = 'PUSH\_CLICK' THEN COUNT (\*) ELSE 0 END PUSH\_CLICK,  CASE WHEN ACTION\_TYPE\_MAJOR = 'PUSH\_VIEW' THEN COUNT (\*) ELSE 0 END PUSH\_VIEW,  CASE WHEN ACTION\_TYPE\_MAJOR = 'V' THEN COUNT (\*) ELSE 0 END PAGE\_VIEW  FROM [dbo].[User\_Activity] AS U1  GROUP BY DATEPART(HOUR, action\_request\_time),ACTION\_TYPE\_MAJOR  ) AS T1  GROUP BY T1.[Hours]  ; |

1. Data visualization

* Number of action records of each major type for per hour.

|  |
| --- |
| No description available. |