

World of Tanks Data Analyst Test Case

This is a test task from the World of Tanks Product Analytics team. All tasks are grouped into 3 sections: Data Mining, Analysis Requirements Formalization, and Analysis Report Preparation. These are examples of certain stages of our department's work.

When completing tasks, you can use any information you receive from this document or any other open sources of information about World of Tanks. Also, we do not limit the candidates in terms of the methods and technologies used to solve the tasks assigned to them—use any method that is convenient for you to achieve the result of the task. Good luck!

Task I—Data Collection and Processing

When solving tasks from this section, you can use any implementation of the SQL type (Oracle, Impala/Hive, MySQL, etc.) Be sure to specify which syntax you are using.

Task #1.1: The database contains the `player_battles` table that contains data about battles played by each player during the day:

player_id	battle_id	vehicle_name	dt	exp
10324	1563321487	ussr:A96_Sherman	14.09.2018	60
10226	1745949855	uk:GB72_Conqueror	14.09.2018	565
10275	1745959321	ussr:R3_BT-7	15.09.2018	0
10276	1660434033	japan:J24_To_130	15.09.2018	309
10275	1660434033	germany:G47_VK3001P	15.09.2018	85
...

where:

- *player_id*: the unique identifier of a player.
 - *battle_id*: the unique battle identifier that is common for all participants in one battle.
 - *vehicle_name*: the name of the vehicle in which the player entered the battle. Marks like 'ussr', 'uk', 'germany' are identifiers of the vehicle nation.
 - *dt*: battle date.
 - *exp*: the amount of experience received by the player in a battle.
- 1) Please write an SQL-script that allows to display the list of vehicle nations that have fought more than 10,000 battles as of the selected date, for example, 15.09.2018.
 - 2) Please write an SQL script that allows you to get the information about the number of unique players who fought a certain number of battles by cohorts. The cohorts should be as follows:
 - [0;10) battles
 - [10;20) battles
 - [20;50) battles
 - [50;100) battles
 - [100; +inf) battles
 - 3) Please write an SQL script that allows you to count the number of players which were active on both days: 14.09.2018 and 15.09.2018.



Task #1.2*: You have two tables (see below) which one of them contains battle statistics per player and the second one contains a list of unique identifiers of players who are developers (*dev_player_id*):

battle_id	battle_start	player_id	team_id	tank_id	damage	credits	exp	team_won_id
1745949855	3.07.22 06:22	980985045322345	1	32607	595	668	721	1
1745949855	3.07.22 06:22	980985045322349	2	90020	269	688	758	1
1660434033	23.12.22 16:28	980985045322355	1	24461	654	913	955	2
1660434036	23.12.22 16:28	980985045322356	2	27809	863	952	509	2
...
1745959321	8.03.23 13:54	980985045322353	1	12472	80	261	953	0
1745959321	8.03.23 13:54	980985045322356	2	73568	799	860	555	0
1660434033	8.03.23 13:54	980985045322357	1	38822	999	709	538	0
1660434378	9.08.23 23:02	980985045322356	1	97796	363	798	181	1
1745959323	9.09.22 23:02	980985045322349	2	97018	954	416	398	2

where:

- *battle_id* : the unique battle identifier which is common for all participants in one battle;
- *battle_start* : the exact time when a battle started;
- *player_id* : the unique identifier of a player;
- *team_id* : the identifier of a team in which a player was distributed in a battle. There are two teams in a battle;
- *tank_id* : the unique identifier of a vehicle in which the player entered the battle;
- *damage* : the amount of damage done to the enemies by the player in a battle;
- *credits* : the number of credits received by the player in a battle;
- *exp* : the amount of experience received by the player in a battle;
- *team_won_id* : the unique identifier of the team who won a battle.

dev_player_id
1660434033
1660434378
...

Using the tables above, please write an SQL script to gather following metrics for each player the following:

- the amount of damage done to the enemies and earned credits in the last player's won battle excluding developers;
- date and time of the first lost player's battle excluding developers;



Task II—Analysis Requirements Formalization

In World of Tanks there are launched the seasons of Battle pass with the duration around 3 months each ([Battle pass description in the WoT Portal](#)).

The Producer came to you as a Data Analyst and requested to provide the responsible team with monitoring of the Battle pass health metrics to understand the current state of activity and, if necessary, make prompt decisions in the World of Tanks project. Please, complete the following:

- Formulate 10 questions that you would ask the Producer to better understand the in-game activity that is being launched.
- Suggest at least 10 metrics that you would suggest to the Producer to monitor the event status.
- Suggest ways and tools to complete the received request.

* Additionally (optional): Describe the stages/sequence of your actions to complete the request.



Task III—Analysis Report Preparation

Task 3.1

The attached *data_1.csv* file contains battles data for two groups of players divided by availability of some changes in setting and configuration in the game. For group A changes were available, for group B were not. Data includes activity from **2022-10-01** till **2022-11-29**. From **2022-10-27** to **2022-11-29** changes were available for Group A.

Field	Description
player_id	the unique identifier of a player
battle_id	the unique battle identifier which is common for all participants in one battle
dt	date at which battle happened
player_group	group of player (doesn't change within period)
in_battle_presence_time	Time in seconds which player spend in a certain battle
damage_dealt	damage that player dealt to other vehicles during this battle
kills_made	number of killed tanks that player made during this battle
vehicle_lvl	level of player's vehicle used in a certain battle
account_created_at	date and time when a player created account in the game

Please, show the approaches in data analysis and visualization known to you. Feel free use any tools, technics, and approaches to complete the task:

- Examine the data set and carry out the analysis. We expect some report on your research that would show the performance of released changes in the game and other interesting facts that you're able to find in the data set. It's supposed to be the report for business stakeholders who make decision about launched improvements.
- Based on the results of the generated report, briefly describe the observations and conclusions that you made when processing and analyzing the data.

**Optional: Based on the results, provide the Product Manager with recommendations on how to improve the performance of updates and new vehicle branches in the future.*

Task 3.2

In the attached *data_2.csv* file, there are data on the number of active players during the period of 2016–2017, where:

Field	Description
DAY_ID	Date
DAU	Daily Active Users: the number of unique players active on the reporting day.

Make a forecast on the DAU index for the period of 3 months of 2018 (January–March). Please specify the method and approach used to build the predicted values.

