SQL TEST

The following corresponds to an extract of 4 Onepark database tables. Analyze them and answer the questions:

Table 1. BOOKINGS

id	purchase_id	state	created_at	begin_at	end_at	park_id	site
1024924	1024977	finished	2019-02-13	2019-02-21	2019-02-26	3879	FR
1020088	1020141	error	2019-02-10	2019-02-20	2019-03-07	2932	FR
960281	960333	finished	2019-01-03	2019-01-18	2019-01-21	4282	ES
1007066	1007119	confirmed	2019-02-03	2019-03-02	2019-03-12	51	FR
992173	992226	finished	2019-01-24	2019-02-10	2019-02-19	2782	FR
986868	986921	confirmed	2019-01-20	2019-03-12	2019-03-14	3024	FR
970333	970385	finished	2019-01-09	2019-02-01	2019-02-04	4282	ES
962774	962826	finished	2019-01-04	2019-01-04	2019-01-10	3799	ES
1023374	1023427	confirmed	2019-02-12	2019-04-10	2019-04-18	2950	FR
1005949	1006002	finished	2019-02-02	2019-02-02	2019-02-04	2882	FR
1044052	1044105	started	2019-02-24	2019-02-24	2019-02-27	3585	FR
997307	997360	error	2019-01-27	2019-02-16	2019-03-04	4412	FR
1046075	1046128	confirmed	2019-02-25	2019-03-01	2019-03-05	3572	FR
1040834	1040887	confirmed	2019-02-22	2019-06-28	2019-07-06	4412	FR
1002315	1002368	cancelled	2019-01-30	2019-02-02	2019-02-08	3275	FR
959768	959820	finished	2019-01-02	2019-01-05	2019-01-07	2792	FR
1022029	1022082	confirmed	2019-02-11	2019-04-04	2019-04-05	2671	FR
1006724	1006777	finished	2019-02-02	2019-02-04	2019-02-24	4412	FR
1019372	1019425	finished	2019-02-10	2019-02-26	2019-02-28	3577	FR

Table 2. PARKS

id	created_at	address_id	market_segment	Type	default_capacity
3214	2016-06-29	11717	city_center	Hotel	50
4436	2018-11-05	14985	airport	Hotel	100
4356	2018-09-12	14772	airport	Discounter	20
4457	2018-11-16	15041	airport	Official	30
4412	2018-10-10	14919	airport	Official	50
2731	2015-08-09	8370	city_center	Hotel	25
4398	2018-09-28	14878	airport	Discounter	5
4322	2018-08-20	14681	airport	Hotel	20
2896	2015-10-12	10913	station	PPI	5
3645	2017-07-21	12784	city_center	PPI	20
3016	2016-01-07	11180	station	Official	5
3234	2016-07-20	11762	event	PPO	10
3955	2018-02-05	13636	station	PPI	10
4361	2018-09-18	14786	airport	Hotel	20

Table 3. PURCHASES

id	user_id	price	discount_price
1031371	691898	10	0
1038616	315693	12	0
1000703	773917	54.8	0
979410	774310	13.7	0
1046409	691898	11	1.1
957771	691898	56.7	0
966344	766553	95	0
1004734	790361	54	0
1016651	451556	30	0
963982	761091	18	0
1043243	815591	74.4	0
1045297	816989	21.25	0
1013198	722770	46	0
1030471	806935	76.95	0
1021502	801076	55	0
966787	745312	74	0

Table 4. USERS

id	role	created_at	zip_code
691898	customer	2015-12-20	7784
315693	customer	2014-10-19	75015
773917	customer	2015-11-02	3000
691898	customer	2018-01-03	70230
817709	customer	2018-06-25	1671
761097	customer	2018-11-30	76135
691898	customer	2018-06-15	45340
790361	customer	2018-07-31	44980
451556	customer	2018-01-01	6560
761091	customer	2018-03-20	92000
815591	customer	2017-04-11	35430
816989	customer	2017-09-18	76290
691898	customer	2017-03-10	75017
806935	customer	2016-01-01	1806
801076	customer	2015-10-01	71700
745312	customer	2014-04-24	23889

Questions :

1/ Write an SQL query to find **the number of bookings by carpark** made in **January 2019**, as well as the **average stay duration** (DMS). The output should has 3 columns named: “park id”, “Nb bookings”, and “DMS”. (10%)

2/ Write a query to find the **revenue per month** generated from the “airport” segment carparks, as well as the revenue **growth rates** compared to the previous **month** and compared to the previous **year**. The output should consist of 4 columns named "Month", "Turnover", "MoM growth" and "YoY growth". (10%)

3/ Write a query to find the **repeat rate** for airport customers (i.e. the percentage of customers who have made at least a second booking). (20%)

4/ Write a query to find the **average time between 2 reservations (in days) for the same customer.** (25%)

5/ Write a query to find out the percentage of customers who, before making a 2nd reservation in an official carpark (“returning customers”), have made a reservation in: 1) a hotel, and 2) a PPI (independent public carpark). In other words, identify the **origin of the reservation that was just prior to a reservation made in an official carpark**, per customer. How many clients have booked in a hotel just before booking in an official carpark? How many clients have booked in a PPI just before booking in an official carpark (number of clients who have booked in a hotel and in a PPI divided by the number of “returning” clients in an official carpark). (30%)

6/ When trying to solve an analysis that you are not familiar with (complex SQL query), what is your strategy to solve the porblem? (5%)

Bonus: if you have additional time, please answer to the following questions

7/ Write a query to find the **maximum occupancy rate** of carpark # 4412 over the last 15 days (i.e. the maximum number of spaces occupied on a day, divided by the carpark capacity).

8/ What is the purpose of creating a Data Warehouse within a company?

9/ What are the criterias that you consider when choosing a chart type to visualize the results of an analysis, to have an efficient data visualization?