

# **PATIENTS' APPOINTMENT ANALYSIS OF VIRGINIA STATE OF USA**

## **1.Introduction**

The main objective of this project is to help the clinic reduce missed appointments , enhance communication with patients, and improvement in resource allocation by identifying and analyzing key patterns using SQL.

## **2. Dataset Overview**

The dataset contain real-world medical appointment records from **Virginia**, a state in the United States. The data contains information on 9916 appointments, including patient demographics, clinical details, and attendance rates such as :

PatientId ,AppointmentID, Gender, ScheduledDay,AppointmentDay, Age and their Neighbourhood.It also contain Scholarship that indicates whether the patient is enrolled in welfare program or not, Some health indicators such as Hypertension, Diabetes, Alcoholism, Handcap are also present in dataset. List of number of SMS\_received as appointment reminder ,Date.diff which indicates the calculated difference in days between the scheduled and appointment date,List of number of patients who actually Showed\_up on appointment day and AppointmentStatus.

## **3. Methodology and SQL Tasks**

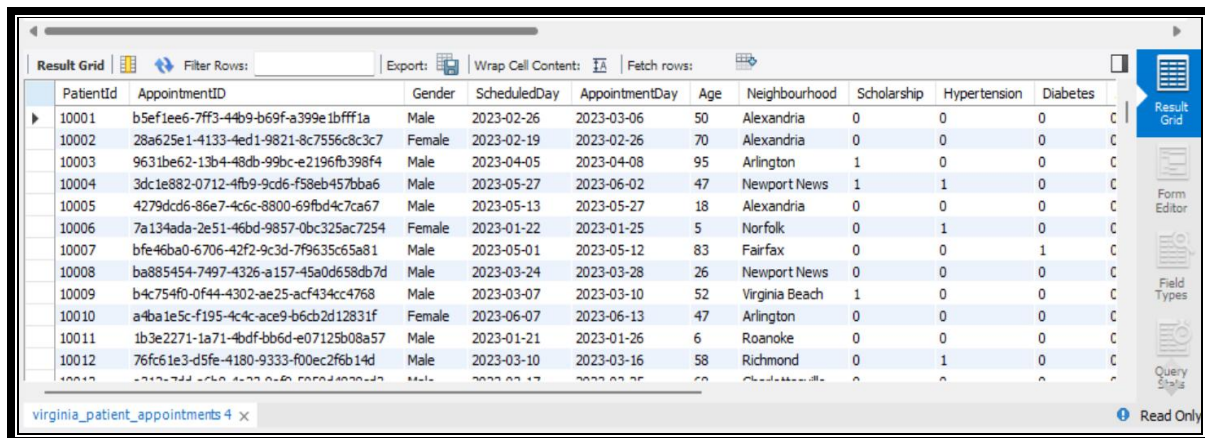
## Basic SQL & Data Retrieval

### 1- Retrieve all columns from the Appointments table.

#### Code

```
-- Q1.Retrieve all columns from the Appointments table.  
select * from virginia_patient_appointments; # To see Complete record
```

#### Output



The screenshot shows a database interface with a 'Result Grid' tab selected. The grid displays 12 rows of data from the 'virginia\_patient\_appointments' table. The columns are: PatientID, AppointmentID, Gender, ScheduledDay, AppointmentDay, Age, Neighbourhood, Scholarship, Hypertension, and Diabetes. The data is as follows:

PatientID	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hypertension	Diabetes
10001	b5ef1ee6-7ff3-44b9-b69f-a399e1bfff1a	Male	2023-02-26	2023-03-06	50	Alexandria	0	0	0
10002	28a625e1-4133-4ed1-9821-8c7556c8c3c7	Female	2023-02-19	2023-02-26	70	Alexandria	0	0	0
10003	9631be62-13b4-48db-99bc-e2196fb398f4	Male	2023-04-05	2023-04-08	95	Arlington	1	0	0
10004	3dc1e882-0712-4fb9-9cd6-f58eb457bba6	Male	2023-05-27	2023-06-02	47	Newport News	1	1	0
10005	4279dcd6-86e7-4c6c-8800-69fbd4c7ca67	Male	2023-05-13	2023-05-27	18	Alexandria	0	0	0
10006	7a134ada-2e51-46bd-9857-0bc325ac7254	Female	2023-01-22	2023-01-25	5	Norfolk	0	1	0
10007	bfe46ba0-6706-42f2-9c3d-7f96335c65a81	Male	2023-05-01	2023-05-12	83	Fairfax	0	0	1
10008	ba885454-7497-4326-a157-45a0d658db7d	Male	2023-03-24	2023-03-28	26	Newport News	0	0	0
10009	b4c754f0-0f44-4302-ae25-acf434cc4768	Male	2023-03-07	2023-03-10	52	Virginia Beach	1	0	0
10010	a4ba1e5c-f195-4c4c-ace9-b6cb2d12831f	Female	2023-06-07	2023-06-13	47	Arlington	0	0	0
10011	1b3e2271-1a71-4bdf-bb6d-e07125b08a57	Male	2023-01-21	2023-01-26	6	Roanoke	0	0	0
10012	76fc61e3-d5fe-4180-9333-f00ec2f6b14d	Male	2023-03-10	2023-03-16	58	Richmond	0	1	0

All records of Appointments data

### 2- List the first 10 appointments where the patient is older than 60.

## Code

```
-- Q2. List the first 10 appointments where the patient is older than 60.
select * from
virginia_patient_appointments
where age >60
order by age limit 10;
```

## Output

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

Fetch rows:

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hypertension	Diabetes
▶	19797	7ea3dbf3-d052-4f1f-9e31-a2c8dd51e00b	Male	2023-01-22	2023-02-01	61	Norfolk	0	0	0
	19323	3783020d-2778-49a1-80f9-8fb99464e261	Male	2023-04-24	2023-04-27	61	Richmond	0	0	0
	19785	b6ce2a8c-d243-40c5-ae4f-53c6509629fc	Male	2023-06-07	2023-06-18	61	Norfolk	1	0	0
	19553	7f1a8f72-a0f3-47e9-b3a4-9aceda435125	Female	2023-06-23	2023-06-29	61	Richmond	0	1	0
	19347	96d4beb8-7cc7-4e77-b518-3180c325d7ee	Male	2023-02-12	2023-02-25	61	Roanoke	0	0	1
	19841	e9dbdbca-4621-4fd8-b828-bfe864b23834	Female	2023-06-18	2023-06-22	61	Newport News	0	0	1
	19613	230c1c0d-1628-4bf4-9d9b-3ca44c0d4017	Male	2023-02-07	2023-02-18	61	Arlington	0	0	0
	19415	53f65d85-f198-4abd-927a-08aa1ccd29c6	Male	2023-01-20	2023-01-28	61	Roanoke	0	1	0
	19702	a2aa7f9a-79fe-4183-93f5-4676e2572729	Female	2023-06-29	2023-07-01	61	Charlottesville	0	0	0
	14365	cf1af575-24b6-491c-9df1-f49be59a3a64	Male	2023-05-17	2023-05-28	61	Roanoke	0	0	0

First 10 records of Patients older than 60

## 3-Show the unique neighborhoods from which patients came.

## Code

```
-- Q3. Show the unique neighborhoods from which patients came
• select
distinct Neighbourhood
as Unique_Neighbourhood
from virginia_patient_appointments;
```

## Output

	Unique_Neighbourhood
▶	Alexandria
	Arlington
	Newport News
	Norfolk
	Fairfax
	Virginia Beach
	Roanoke
	Richmond
	Charlottesville
	Chesapeake

### Patients' Unique Neighbourhoods (Area of living)

The above output shows that there are total ten areas from where people came for appointment booking.

#### 4. Find all female patients who received an SMS reminder. Give count of them

##### Code# 1 (All female patients who received SMS reminder for appointment)

```
# female patients_ids who received sms reminder
• select PatientId,Gender,SMS_received
  from virginia_patient_appointments
 where SMS_received=1 and Gender="Female"
 group by PatientId,Gender,SMS_received; # only female patients with sms reminder will get displayed
```

## Output

	PatientId	Gender	SMS_received
▶	10002	Female	1
	10010	Female	1
	10019	Female	1
	10021	Female	1
	10023	Female	1
	10025	Female	1
	10026	Female	1
	10029	Female	1
	10033	Female	1
	10034	Female	1
	10035	Female	1
	10037	Female	1
	10043	Female	1
	10044	Female	1

virginia\_patient\_appointments 7 ×

Output

Female Patients who received an SMS-Reminder for their Appointment

## Code#2 (Total count of female patients who received SMS reminder)

```
• select distinct count(*) as total_count
  from virginia_patient_appointments
 where Gender="Female" and SMS_received=1; # total count of female patients who get sms reminder will be displayed
```

## Output

	total_count
▶	3465

Total no. of female pateints who received an SMS Reminder for their appointment

## 5. Display all appointments scheduled on or after '2023-05-01' and before '2023-06-01'.

It means list of all appointments scheduled from 1<sup>st</sup> may 2023 onward,before 1<sup>st</sup> june 2023.

## Code

```
# STEP-1 OF Q(5): First we need to modify text-date to its proper 'date format' for further operation

• set sql_safe_updates=0;
• update virginia_patient_appointments
  set ScheduledDay= str_to_date(ScheduledDay,"%m/%d/%Y"); #DATE MODIFICATION
• alter table virginia_patient_appointments
  modify ScheduledDay date;

# STEP-2 OF Q(5): All appointments scheduled on or after '2023-05-01' and before '2023-06-01'.
• select *,
  case
  when
    ScheduledDay between '2023-05-01' and '2023-05-31' then 1
  end as scheduled_appointments
  from virginia_patient_appointments
  having scheduled_appointments is not null;
```

## Output

PatentId	AppointmentId	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hypertension	Diabetes
10004	3dc1e882-0712-4fb9-9cd6-f58eb457bba6	Male	2023-05-27	2023-06-02	47	Newport News	1	1	0
10005	4279dcd6-86e7-4c6c-8800-69fbd4c7ca67	Male	2023-05-13	2023-05-27	18	Alexandria	0	0	0
10007	bfe46ba0-6706-42f2-9c3d-7f9635c65a81	Male	2023-05-01	2023-05-12	83	Fairfax	0	0	1
10031	124abc16-de4c-4944-8678-4f498bdeb84a	Male	2023-05-27	2023-06-09	20	Roanoke	0	0	1
10033	374e3e12-ebed-45d4-84f2-99efc26cd636	Female	2023-05-10	2023-05-17	37	Charlottesville	0	0	0
10035	e31896d5-7216-406c-b1ab-2c49ec817e3e	Female	2023-05-14	2023-05-25	87	Richmond	0	1	0
10039	75766872-d5cb-40a3-a9cb-fc4e2f4bc0ce	Male	2023-05-11	2023-05-25	9	Chesapeake	1	0	0
10045	8a38ef1d-015d-441f-8e22-0c2dea771919	Female	2023-05-19	2023-05-26	16	Newport News	0	1	1
10061	a5588005-6527-4d38-b3db-22db6b76c990	Male	2023-05-05	2023-05-05	18	Arlington	0	0	0
10066	5fa686f0-961d-4530-a068-f9edbf9209f	Female	2023-05-30	2023-06-03	97	Roanoke	0	0	0
10079	478cd79d-0856-4c4b-9ea4-6e7d3cfa73f8	Male	2023-05-27	2023-05-28	99	Alexandria	0	1	0
10082	f505b662-5937-4710-a829-1eebe614a0d9	Female	2023-05-24	2023-06-05	83	Virginia Beach	0	0	0

All Appointments scheduled on or after 1<sup>st</sup> May 2023 but before 1<sup>st</sup> June 2023

## Data Modification & Filtering

6. Update the 'Showed\_up' status to 'Yes' where it is null or empty

## Code

```
-- STEP(1) OF Q(6)
• select Showed_up from virginia_patient_appointments
  where Showed_up is NULL; # check if there is a missing value is present or not

-- STEP(2) OF Q(6)
• set sql_safe_updates=0;
• update virginia_patient_appointments
  set Showed_up = 'Yes'
  where
    Showed_up is NULL;
• select * from virginia_patient_appointments; # check if missing value is still present or not
```

## Output

	Age	Neighbourhood	Scholarship	Hypertension	Diabetes	Alcoholism	Handcap	SMS_received	Date.diff	Showed_up
	94	Charlottesville	0	0	0	0	0	1	7	Yes
	11	Arlington	0	1	0	0	1	1	1	Yes
	3	Roanoke	0	1	0	0	0	1	0	No
	41	Chesapeake	1	0	1	0	0	1	1	Yes
	12	Charlottesville	0	0	0	0	0	0	5	Yes
	97	Fairfax	1	0	0	1	0	1	5	Yes
	84	Newport News	0	0	0	0	0	0	14	No
	85	Virginia Beach	0	0	0	0	0	1	6	Yes
	3	Fairfax	0	0	0	0	0	1	11	Yes
	81	Virginia Beach	0	1	0	0	0	1	9	Yes
	68	Newport News	0	1	0	0	0	1	14	No
	18	Arlington	0	0	0	0	0	1	0	Yes
	1	Richmond	1	0	0	0	0	1	7	Yes
	11	Norfolk	1	1	0	0	0	1	4	Yes
	80	Alexandria	1	0	0	0	0	1	3	Yes

Updated Record showing **Showed\_up=yes** on empty/Null values

This update indicates that the patients who appeared for appointment are Shown as 'Yes' and those who didn't appear are indicated by 'No'. This thing is helpful to know how many patient couldn't appear and then appointment attendance can be improved by any communication or any other channel.

## 7. Add a new column AppointmentStatus using a CASE statement:

○ 'No Show' if Showed\_up = 'No'

○ 'Attended' otherwise

### Code

The purpose is to see how much the appointment attendance is good or bad,so we can find reason for bad attendance further.

```
# PART (1) OF Q(7)
• alter table virginia_patient_appointments
  add column AppointmentStatus varchar(20); #adding a column named AppointmentStatus

# PART (2) OF Q(7)
• set sql_safe_updates=0;
• update virginia_patient_appointments
  set AppointmentStatus = (select
    case when Showed_up = 'No' then 'No Show'
    else 'Attended'
  end ); #update according to given condition

• select * from virginia_patient_appointments; #check modification in complete table
```

### Output



Result Grid												
Filter Rows:												
Export:												
Wrap Cell Content:												
Fetch rows:												
Age	Neighbourhood	Scholarship	Hypertension	Diabetes	Alcoholism	Handcap	SMS_received	Date.diff	Showed_up	AppointmentStatus	days	
8	Chesapeake	0	1	0	0	0	0	3	No	No Show	Thursday	
14	Newport News	0	1	1	0	0	1	12	Yes	Attended	Tuesday	
2	Fairfax	0	1	0	1	0	0	10	Yes	Attended	Thursday	
92	Chesapeake	0	0	0	0	0	1	11	Yes	Attended	Thursday	
69	Richmond	0	1	0	0	0	1	7	Yes	Attended	Thursday	
16	Roanoke	0	1	0	0	0	0	7	Yes	Attended	Thursday	
25	Charlottesville	1	1	1	0	0	0	13	Yes	Attended	Saturday	
20	Chesapeake	0	0	0	0	0	1	5	Yes	Attended	Tuesday	
71	Roanoke	0	1	0	0	0	0	14	No	No Show	Friday	
20	Roanoke	0	0	1	0	0	1	13	No	No Show	Friday	
10	Alexandria	1	0	0	0	0	0	8	Yes	Attended	Tuesday	
37	Charlottesville	0	0	0	0	0	1	7	No	No Show	Wednesday	
71	Norfolk	0	1	0	0	0	1	7	No	No Show	Thursday	
87	Richmond	0	1	0	0	0	1	11	No	No Show	Thursday	
34	Virginia Beach	1	1	0	0	0	0	4	Yes	Attended	Tuesday	
12	Newport News	1	0	0	0	0	1	5	Yes	Attended	Saturday	

Updated Record showing AppointmentStatus

## 8. Filter appointments for diabetic patients with hypertension.

The purpose is to see patients who have diagnosed with diabetes and hypertension. we can see which patients are having health indicators of diabetes and hypertension and later we can also find their total number as well for further insights.

### Code

```

• Select AppointmentID, PatientId, Diabetes, Hypertension
from virginia_patient_appointments
where Diabetes=1 and
Hypertension=1; # patients who have diabetes and hypertension will be showed

```

### Output

	AppointmentID	PatientId	Diabetes	Hypertension
▶	ce74ec56-db42-4295-95a2-b7e4351399ed	10023	1	1
	2f33d653-4421-4249-a3f6-b904fab34b7f	10028	1	1
	8a38ef1d-015d-441f-8e22-0c2dea771919	10045	1	1
	edfc1ab2-c198-4cfb-a5cf-9236bdce8c96	10047	1	1
	cf03c9ad-72d5-48d4-9ed6-d4366cfc3638	10128	1	1
	2c01b5a6-c9dd-40a8-a78f-5465228dc422	10168	1	1
	d9e07cbd-1734-4957-8341-1e59a5680f62	10184	1	1
	6c74eed4-4578-44e6-b32b-ad052c682589	10216	1	1
	f9848030-9fa9-45b1-ae18-76034ad551ad	10234	1	1
	a5c5677f-d8af-426c-bd1d-53ca4e027022	10268	1	1
	fcc6febe-7b48-4fd1-8185-361480f377c2	10314	1	1
	0c44c0fd-e13c-464d-b23e-860e0ae35494	10378	1	1
	f38f922c-56c0-472f-b747-e5cd94571290	10390	1	1
	14cb07a9-0180-4ceb-a665-c3027c0ba9d2	10409	1	1
	ee31bc55-a973-48f7-8f7d-59dc4b9486e6	10411	1	1
	ce8a4293-673a-48f7-a193-ab0809d73a75	10440	1	1
	4eaeba0b-f07a-4ab7-8b9f-82a53f76f3b0	10514	1	1

Appointments for Diabetic and Hypertension Health Indicators

**9- Order the records by Age in descending order and show only the top 5 oldest patients.**

This is helpful in finding maximum age group interested in appointment with doctor,so we can improve our required facilities according to age group in order to maintain and improve patient retention like comfortable chairs,increased number of helpers in waiting area for older people for their support and counselling sessions.

## Code

```
• Select * from  
virginia_patient_appointments  
order by age desc limit 5; #show only the top 5 oldest patients in descending order
```

## Output

PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hypertension	Diabetes
19689	92d1e32e-8d0d-4cb9-b4bf-f3cdfc6fa36b	Male	2023-02-03	2023-02-07	99	Fairfax	0	0	1
19876	9e8f862b-537e-474e-9b5c-aace44fcc9f4	Female	2023-05-11	2023-05-25	99	Arlington	0	0	0
19591	4a1c3baa-d25c-4f82-97b3-da76bfa80526	Female	2023-04-06	2023-04-17	99	Alexandria	0	1	0
19578	5ed5e6e7-035d-4518-af0c-7e2d39441fc4	Male	2023-03-25	2023-03-25	99	Roanoke	1	0	0
19711	05eb7c54-7f1a-4298-870a-6e4886a05e8e	Male	2023-04-01	2023-04-08	99	Roanoke	0	0	1

Top 5 oldest Patients' Record in Descending order

### 10- . Limit results to the first 5 appointments for patients under age 18.

The Insights about age group of children under 18 are helpful so we can involve more counselling session as the young children are more exposed to anxiety and boredom as compared with other adults. So, additional mental health sessions according to their convenience can be a good initiative for their better attendance in appointment.

we can perform testing initially on first 5 under 18 children and then decide for other children as well.

## Code

```
• Select * from virginia_patient_appointments
  where age<18
  order by age asc limit 5;
```

## Output

Result Grid										
Filter Rows:		Export:		Wrap Cell Content:		Fetch rows:				
PatientID	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hypertension	Diabetes	
17876	b4c6b8e1-0fcc-4605-ba16-c195b69859d7	Female	2023-06-20	2023-07-01	1	Norfolk	0	0	0	
11985	988ba37a-0829-4bfd-b881-d322fe994342	Female	2023-03-03	2023-03-06	1	Alexandria	0	0	0	
17786	c7de1e99-4b34-4a3f-ba96-c12c354686f3	Male	2023-06-01	2023-06-04	1	Richmond	0	0	0	
16797	615cd51b-6083-4ecd-af32-5b13caf21dec	Male	2023-06-09	2023-06-10	1	Fairfax	0	1	1	
13989	a56b22a2-3b6a-4a85-b765-2bf7cb286bab	Male	2023-05-29	2023-06-12	1	Chesapeake	0	1	1	

First 5 Appointments for patients under age 18

## Aggregation & Grouping

11. Find the average age of patients for each gender.

## Code

```
-- Q11. Find the average age of patients for each gender.

• Select Gender, round(avg(age),2) from virginia_patient_appointments
  group by Gender;
```

## Output

	Gender	round(avg(age),2)
▶	Male	50.36
	Female	49.74

Average age of patients for each gender

## 12. Count how many patients received SMS reminders, grouped by Showed\_up status

### Code

```
• select SMS_received,  
  Showed_up,count(*) as patient_count  
  from virginia_patient_appointments  
  group by Showed_up,SMS_received  
  having SMS_received=1; # patients who received SMS-reminder will get display
```

## Output

	SMS_received	Showed_up	patient_count
▶	1	Yes	5403
	1	No	1478

Total Patients who received SMS reminder

## 13- Count no-show appointments in each neighborhood using GROUP BY

### Code

```
select count(AppointmentStatus="No Show") as No_show_count, Neighbourhood  
from virginia_patient_appointments  
group by Neighbourhood;
```

## Output

	No_show_count	Neighbourhood
▶	1018	Alexandria
	1027	Arlington
	991	Newport News
	999	Norfolk
	977	Fairfax
	946	Virginia Beach
	980	Roanoke
	1014	Richmond
	956	Charlottesville
	1008	Chesapeake

Total missed Appointment missed in each neighbourhood

14. Show neighborhoods with more than 100 total appointments (HAVING clause).

## Code

```
-- Q14. Show neighborhoods with more than 100 total appointments (HAVING clause).
• select Neighbourhood,count(AppointmentID) as total_appointments_count
  from virginia_patient_appointments
  group by Neighbourhood
  having total_appointments_count>100;
```

## Output

	Neighbourhood	total_appointments_count
▶	Alexandria	1018
	Arlington	1027
	Newport News	991
	Norfolk	999
	Fairfax	977
	Virginia Beach	946
	Roanoke	980
	Richmond	1014
	Charlottesville	956
	Chesapeake	1008

Neighborhoods with more than 100 total appointments

**15. Use CASE to calculate the total number of:**

○ children (Age < 12) ○ adults (Age BETWEEN 12 AND 60) ○ seniors (Age > 60)

**Code**

```
• select
  count(case when Age<12 then 1 end )as children,
  count(case when Age between 12 and 60 then 1 end) as adults,
  count(case when Age >60 then 1 end) as seniors
from virginia_patient_appointments;
```

**Output**

	children	adults	seniors
▶	1074	4910	3932

Total number of Children,Adults and seniors for appointment

## **Window Functions**

**16. Tracks how appointments accumulate over time in each neighbourhood. (Running**

**Total of Appointments per Day)**

It is helpful in finding the varying trends of appointment of each day.

## Code

```
• select Neighbourhood,AppointmentDay,  
count(AppointmentID) as Total_Appointments_per_day,  
sum(count(AppointmentID) ) over (partition by Neighbourhood order by AppointmentDay) as Running_total_Appointments  
from virginia_patient_appointments  
group by Neighbourhood,AppointmentDay  
order by Neighbourhood ;
```

## Output

	Neighbourhood	AppointmentDay	Total_Appointments_per_day	Running_total_Appointments
▶	Alexandria	2023-01-03	1	1
	Alexandria	2023-01-05	1	2
	Alexandria	2023-01-06	1	3
	Alexandria	2023-01-07	5	8
	Alexandria	2023-01-08	3	11
	Alexandria	2023-01-09	3	14
	Alexandria	2023-01-10	5	19
	Alexandria	2023-01-11	6	25
	Alexandria	2023-01-12	5	30
	Alexandria	2023-01-13	2	32
	Alexandria	2023-01-14	5	37
	Alexandria	2023-01-15	2	39
	Alexandria	2023-01-16	2	41
	Alexandria	2023-01-17	3	44
	Alexandria	2023-01-18	8	52
	Alexandria	2023-01-19	5	57

Result 26 x

Running total of Appointment per day

### 17. Use Dense\_Rank() to rank patients by age within each gender group.


The insights about ranking are helpful to find highest age and lowest age in each gender and conduct appointments according to their availability hours for better attendance



## Code

```
• select * from virginia_patient_appointments;  
#Use Dense_Rank() to rank patients by age within each gender group.  
• select PatientId, Age, Gender,  
dense_rank() over ( partition by gender order by age desc) as age_dense_rank  
from virginia_patient_appointments  
order by age desc;
```

## Output



PatientId	Age	Gender	age_dense_rank
11155	99	Female	1
11848	99	Female	1
13411	99	Female	1
10546	99	Female	1
10246	99	Female	1
11321	99	Female	1
13468	99	Female	1
13184	99	Female	1
12660	99	Female	1
14162	99	Female	1
11204	99	Female	1
12911	99	Female	1
14877	99	Female	1
13704	99	Female	1
11077	99	Female	1
11141	99	Female	1
13827	99	Female	1

Ranking of patients by age within each gender group

## 18- How many days have passed since last appointment in same neighborhood?

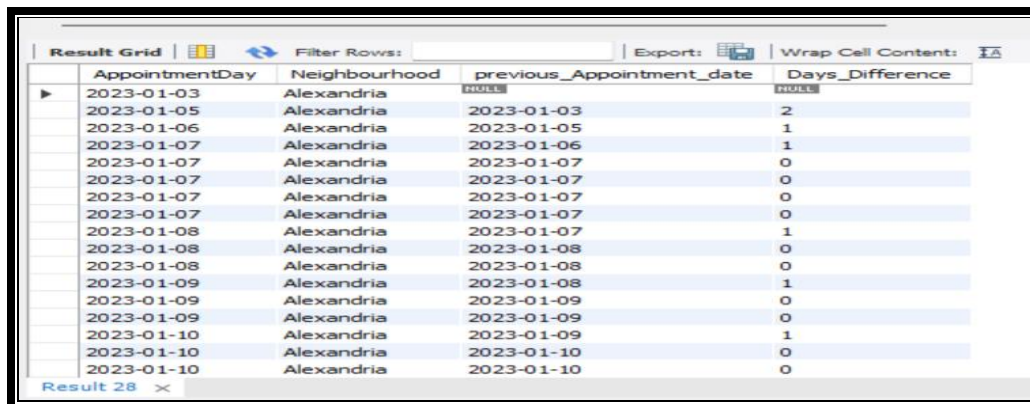
It is helpful for having an estimate about which day is suitable for booking appointment as many people are job person and don't have time in week days usually.

## Code

```
-- Q18 : How many days have passed since the last appointment in the same neighborhood?
-- (Hint:DATEDIFF and Lag) (This helps to see how frequently appointments are happening in each
-- neighborhood.)

• select AppointmentDay,Neighbourhood,
lag(AppointmentDay) over (partition by Neighbourhood order by appointmentDay) as previous_Appointment_date,
datediff(AppointmentDay, lag(AppointmentDay) over (partition by Neighbourhood order by appointmentDay )) as Days_Difference
from virginia_patient_appointments;
```

## Output



The screenshot shows a SQL query result grid with the following columns: AppointmentDay, Neighbourhood, previous\_Appointment\_date, and Days\_Difference. The data is filtered for the Neighbourhood 'Alexandria'. The 'Days\_Difference' column shows the number of days between consecutive appointments. The first row shows a 'miss' for the previous appointment date.

AppointmentDay	Neighbourhood	previous_Appointment_date	Days_Difference
2023-01-03	Alexandria	miss	miss
2023-01-05	Alexandria	2023-01-03	2
2023-01-06	Alexandria	2023-01-05	1
2023-01-07	Alexandria	2023-01-06	1
2023-01-07	Alexandria	2023-01-07	0
2023-01-07	Alexandria	2023-01-07	0
2023-01-07	Alexandria	2023-01-07	0
2023-01-07	Alexandria	2023-01-07	0
2023-01-08	Alexandria	2023-01-07	1
2023-01-08	Alexandria	2023-01-08	0
2023-01-08	Alexandria	2023-01-08	0
2023-01-09	Alexandria	2023-01-08	1
2023-01-09	Alexandria	2023-01-09	0
2023-01-09	Alexandria	2023-01-09	0
2023-01-10	Alexandria	2023-01-09	1
2023-01-10	Alexandria	2023-01-10	0
2023-01-10	Alexandria	2023-01-10	0

No. of days passed since Previous Appointment

**19. Which neighborhoods have the highest number of missed appointments? Use DENSE\_RANK() to rank neighborhoods based on the number of no-show appointments.**

The insights about most and least interested areas in appointments are helpful, so we can improve SMS-reminders and healthcare-awareness campaigns to stimulate people who are not interested.

## Code

```
select Neighbourhood, sum(AppointmentStatus='No Show') as Missed,  
dense_rank() over (order by sum(AppointmentStatus = 'No Show') desc ) AS Missed_Appointment_rank  
from virginia_patient_appointments  
group by neighbourhood;
```

## Output

	Neighbourhood	Missed	Missed_Appointment_rank
▶	Charlottesville	217	1
	Fairfax	215	2
	Roanoke	214	3
	Alexandria	211	4
	Norfolk	211	4
	Chesapeake	210	5
	Arlington	204	6
	Virginia Beach	196	7
	Richmond	193	8
	Newport News	188	9

Neighbourhood “**Charlottesville**” showing Highest Rank of missed appointments

## Complex Query

### 20. Are patients more likely to miss appointments on certain days of the week?

We can first find the days of week from the respective date and then perform further operation on it. So we can divide this thing into two steps which are:

### Code #1 (date format setting and then dayname extraction from appointment day)

```
199 # First we need to modify text date to its proper 'date format' for further operation
200 • set sql_safe_updates=0;
201 • update virginia_patient_appointments
202   set AppointmentDay= str_to_date(AppointmentDay, '%m/%d/%Y'); #DATE FORMAT MODIFICATION
203 • alter table virginia_patient_appointments
204   modify AppointmentDay date;
205
206 # STEP-I OF Q(20) :( DAYS EXTRACTION )
207 • alter table virginia_patient_appointments
208   add column days varchar(20); # column named "days" will be created
209
210 • set sql_safe_updates=0;
211 • update virginia_patient_appointments
212   set days=(SELECT DAYNAME(AppointmentDay) as Days); # column named "days" will be updated by extracting name of days from date
213
214
215 • select * from virginia_patient_appointments; # check whole column to see modifications clearly
216
```

### Code #2 (Extraction of Total count of appointment from dataset and the %age of showed and missed appointment per day)

```
217
218 #STEP -II OF Q(20) 'COUNT OF SCHEDULED APPOINTMENTS', 'COUNT OF SHOWED-UP AND MISSED ' ON EACH DAY
219 -- TOTAL APPOINTMENTS :
220 • select count(*) as total_Appointments from virginia_patient_appointments;
221
222 #count of showed-up, missed, %age of shows and missed according to each day in descending order will be displayed
223 • select Days, sum(case when Showed_up="yes" then 1 end ) as Showed_up,
224   sum(case when Showed_up="No" then 1 end ) as Missed,
225   round(count(case when Showed_up = 'yes' then 1 end ) * 100 / count(*), 2) as percentage_of_shows,
226   round(count(case when Showed_up = 'no' then 1 end ) * 100 / count(*), 2) as percentage_of_no_shows
227   from virginia_patient_appointments
228   group by Days
229   order by percentage_of_no_shows desc;
230
```

## Outputs

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	total_Appointments			
▶	9916			

(Total Appointments in virginia of USA)

	Days	Showed_up	Missed	percentage_of_shows	percentage_of_no_shows
▶	Sunday	1101	316	77.70	22.30
	Monday	1074	291	78.68	21.32
	Saturday	1121	297	79.00	20.93
	Thursday	1180	308	79.30	20.70
	Tuesday	1160	301	79.29	20.57
	Friday	1102	280	79.74	20.26
	Wednesday	1115	266	80.68	19.25

(Total Scheduled Appointments, Total Showed-up & missed Appointment per day)

## 4. Key Findings

- ✓ Total no.of People who booked for appointment are 9916
- ✓ Total no.of female and male patients booked for appointment
- ✓ 1074 children, 4910 adults and 3932 seniors being booked for appointment
- ✓ Neighbourhood of People with most and least interested in appointment
- ✓ Total number of people who didn't showed up even after having SMS reminder are 5403 and 1478 showed up on respective appointment day

- ✓ Certain days of week when people are more likely to miss their appointments were also found out
- ✓ Trends of appointment per day in each neighbourhood were found
- ✓ Neighbourhoods with most and least missed appointments ranks were extracted
- ✓ No of day gaps between each appointments in each neighborhood also found
- ✓ Average age of patients who interested in appointment were also observed
- ✓ Percentage of missed and attended appointment according to each day of week was also calculated from their data.

## **5. Recommendations**

It is recommended to increase healthcare-awareness campaign in neighborhood with high ranks in missed appointments specifically Charolettesville neighbourhood .The SMS reminders must be increased so patients can become more active ,specifically for Sunday ,thursday and Tuesday as missed appointment rate is high in these days.

## **6. Conclusion**

The SQL based analysis of Appointment for Virginia state of USA provides valuable insights about certain days and areas to be focused on,for enhancing attendance of patients.Percentage of missed appointments can be reduced by taking steps on recommended actions provided before,to improve attendance of patients and might be helpful for clinic.

---