# Part 1: Energy Stability and Market Outages

Energy stability is one of the key themes the AEMR management team cares about. To ensure energy security and reliability, AEMR needs to understand the following:

1. What are the most common outage types and how long do they tend to last?
2. How frequently do the outages occur?
3. Are there any energy providers that have more outages than their peers which may indicate that these providers are unreliable?

**Please note that for the whole AEMR Case Study we are only interested in the Outages where the Status = Approved.**

Let's get started!

Question 1.1: In the **AEMR** dataset, write a SQL statement that **COUNTS** the number of **valid (i.e. Status = Approved)** outage events for 2016. This should be grouped by and ordered by the outage **reason**.
The output of your query should have the following column headers:
Please note that we are expecting you to use YEAR() as opposed to EXTRACT() when working with the Start_Time column.
We've included a refresher on the SQL Syntax provided here:
SELECT
FROM
WHERE
GROUP BY
ORDER BY

| Total_Number_Outage_Events | Status | Reason |
|---|---|---|

```sql
-- Your answer here
SELECT COUNT(Status='Approved') AS Total_Number_Outage_Events, Status,
 Reason
FROM AEMR
WHERE Year(start_time) = 2016 AND Status='Approved'
GROUP BY Status, Reason
ORDER BY Reason
```

-------Output-----------

Total_Number_Outage_Events     Status Reason

181     Approved     Consequential

## Task

Question 1.3: Write a SQL statement to **COUNT** the number of **valid** (i.e. Status = Approved) outage events **sorted by their reason** (i.e. Forced, Consequential, Scheduled, Opportunistic)* for 2017.

The output of your query should have the following column headers:

| Total_Number_Outage_Events | Status | Reason |
|---|---|---|
|  |  |  |

1

## Quiz

Question

```
-- Your answer here
SELECT COUNT(Status='Approved') AS Total_Number_Outage_Events, Status,
 Reason
FROM AEMR
WHERE Year(start_time) = 2017 AND Status='Approved'
GROUP BY Status, Reason
ORDER BY Total_Number_Outage_Events
```

-------Output-----------

| Total_Number_Outage_Events | Status | Reason |
|---|---|---|
| 102 | Approved | Opportunistic Maintenance (Planned) |
| 127 | Approved | Consequential |
| 320 | Approved | Scheduled (Planned) |
| 1622 | Approved | Forced |

## Task

Question 1.5: Write a SQL statement that calculates the **average duration in days** rounded to 2 decimal places for each **approved** outage type across both 2016 and 2017. Don't forget to Order this by **Reason** and **Year**.

**Please note**: The average duration in days should be rounded to 2 decimal places for ease of comparison. Additionally, include the overall count of outage events per outage type (i.e. Forced, Consequential, Planned, Opportunistic).

You may find the function TIMESTAMPDIFF() here to be very helpful! Look at the **Unit** argument and think carefully regarding how we might be able to convert this to days.
We've included a hint below to help guide your thinking:
TIMESTAMPDIFF(___) / 60 = Hours
How could we convert this to Days?
**Please note you should not use the unit DAY as this will round up to the nearest day which will result inaccuracies**
The output of your query should have the following column headers:

| Status | Reason | Total_Number_Outage _Events | Average_Outage_Duration_ Time_Days | Year |
|--------|--------|------------------------------|-------------------------------------|------|

```
-- Your answer here
SELECT Status,
       Reason,
       count(reason) AS Total_Number_Outage_Events,
       ROUND(AVG(TIMESTAMPDIFF(minute, start_time, end_time)/1440), 2)
 AS Average_Outage_Duration_Time_Days,
       year(start_time) AS Year
FROM AEMR
WHERE year(start_time) IN('2016', '2017') AND status='Approved'
GROUP BY status, reason, year(start_time)
ORDER BY reason, year
```

| Status | Reason | Total_Number_Outage_Events | Average_Outage_Duration_Time_Days | Year |
|--------|--------|-----------------------------|-----------------------------------|------|
| Approved | Consequential | 181 | 0.36 | 2016 |
| Approved | Consequential | 127 | 0.33 | 2017 |
| Approved | Forced | 1264 | 0.56 | 2016 |
| Approved | Forced | 1622 | 0.79 | 2017 |
| Approved | Opportunistic Maintenance (Planned) | 106 | 0.32 | 2016 |
| Approved | Opportunistic Maintenance (Planned) | 102 | 0.27 | 2017 |
| Approved | Scheduled (Planned) | 380 | 4.57 | 2016 |
| Approved | Scheduled (Planned) | 320 | 5.58 | 2017 |

## 2.How frequently do outages occur?

Question 2.1: Write a SQL statement showing the **monthly COUNT** of all approved outage types (Forced, Consequential, Scheduled, Opportunistic) that occurred for **2016**. Order by **Reason** and **Month**.
The output of your query should have the following column headers:

| Status | Reason | Total_Number_Outage_Events | Month |
|--------|--------|----------------------------|-------|
|        |        |                            |       |

```sql
-- Your answer here
SELECT Status,
       Reason,
       count(Reason) AS Total_Number_Outage_Events,
       month(start_time) AS Month
FROM AEMR
WHERE year(start_time)=2016 AND status='Approved'
GROUP BY status, Reason, month(start_time)
ORDER BY Reason, Month
```

Question 2.2: Write a SQL Statement showing the **monthly COUNT** of all approved outage types (Forced, Consequential, Scheduled, Opportunistic) that occurred during **2017**. Order by **Reason** and **Month**.
The output of your query should have the following column headers:

| Status | Reason | Total_Number_Outage_Events | Month |
|--------|--------|----------------------------|-------|
|        |        |                            |       |

```sql
-- Your answer here
SELECT Status,
       Reason,
       count(Reason) AS Total_Number_Outage_Events,
       month(start_time) AS Month
FROM AEMR
WHERE year(start_time)=2017 AND status='Approved'
GROUP BY status, Reason, month(start_time)
ORDER BY Reason, Month
```

Question 2.2: Write a SQL Statement showing the **monthly COUNT** of all approved outage types (Forced, Consequential, Scheduled, Opportunistic) that occurred during **2017**. Order by **Reason** and **Month**.
The output of your query should have the following column headers:

| Status | Reason | Total_Number_Outage_Events | Month |
|--------|--------|----------------------------|-------|
|        |        |                            |       |

0

```sql
-- Your answer here
SELECT Status,
       Reason,
       count(Reason) AS Total_Number_Outage_Events,
       month(start_time) AS Month
FROM AEMR
WHERE year(start_time)=2017 AND status='Approved'
GROUP BY status, Reason, month(start_time)
ORDER BY Reason, Month
```

Question 2.3: Write a SQL statement showing the total number of all approved outage types (Forced, Consequential, Scheduled, Opportunistic) that occurred for both 2016 and 2017 per month (i.e. 1 – 12). Don't forget to Order this by by **Month** and **Year**.
The output of your query should have the following column headers:

| Status | Total_Number_Outage_Events | Month | Year |
|--------|----------------------------|-------|------|
|        |                            |       |      |

0

```sql
-- Your answer here
SELECT Status,
       count(*) AS Total_Number_Outage_Events,
       month(start_time) AS Month, year(start_time) AS Year
FROM AEMR
WHERE year(start_time) IN (2016,2017) AND status='Approved'
GROUP BY status, Month(start_time), year(start_time)
```

```
ORDER BY Year, Month
```

## 3) Are there any energy providers that have more outages than their peers that may be indicative of being unreliable?

Question 3.1: Write a SQL statement showing the count of all approved outage types (Forced, Consequential, Scheduled, Opportunistic) for all participant codes for 2016 and 2017. Order by **Year** and **Participant_Code**.
The output of your query should have the following column headers:

| Total_Number_Outage_Events | Participant_Code | Status | Year |
|---|---|---|---|
| | | | |

0

```
-- Your answer here
SELECT count(Reason) AS Total_Number_Outage_Events,
       Participant_Code, Status,
       year(start_time) AS Year
FROM AEMR
WHERE year(start_time) IN (2016,2017) AND status='Approved'
GROUP BY status, Participant_Code , year(start_time)
```

Question 3.2: Write a SQL statement showing the **average duration** of all approved outage types (Forced, Consequential, Scheduled, Opportunistic) for all participant codes from 2016 to 2017. Don't forget to order the average duration in descending order with the DESC keyword.
Hints:

- Use **TIMESTAMPDIFF** function to return the time value.

- Please note that the time is less than one day, so you need to calculate **minutes** and convert to **days**. Don't forget to use ROUND() to ensure your answer is rounded to 2 decimal places.

The output of your query should have the following column headers:

| Participant_Code | Status | Year | Average_Outage_Duration_Time_Days |
|---|---|---|---|
| | | | |

```
-- Your answer here
SELECT Participant_Code, Status,
       year(start_time) AS Year,
       ROUND(AVG(TIMESTAMPDIFF(minute, start_time, end_time)/1440), 2)
 AS Average_Outage_Duration_Time_Days
FROM AEMR
WHERE year(start_time) IN (2016,2017) AND status='Approved'
GROUP BY status, Participant_Code , year(start_time)
ORDER BY Year, Average_Outage_Duration_Time_Days DESC;
```

## Part 2: Energy Losses and Market Reliability

When an energy provider provides energy to the market, they are making a commitment to the market and saying; "We will supply X amount of energy to the market under a contractual obligation." However, in a situation where the outages are *forced*, the energy provider intended to provide energy but is unable to provide energy and is forced offline. If many energy providers are forced offline at the same time, it could cause an energy security risk that AEMR needs to mitigate.

To ensure this doesn't happen, the AEMR is interested in exploring the following questions:

1. Of the outage types in 2016 and 2017, what percent were forced outages?
2. What was the average duration for a forced outage during both 2016 and 2017? Have we seen an increase in the average duration of forced outages?
3. Which energy providers tended to be the most unreliable?

Let's get started!

## 1) Of the outage types in 2016 and 2017, what are the respective percentages composed of Forced Outage(s)?

Question 1.1: Write a SQL Statement to COUNT the total number of approved forced outage events for 2016 and 2017. Order by **Reason** and **Year**.
The output of your query should have the following column headers:

| Total_Number_Outage_Events | Reason | Year |
|---|---|---|
| | | |

```
-- Your answer here
SELECT count(*) AS Total_Number_Outage_Events, Reason,
       year(start_time) AS Year
FROM AEMR
WHERE year(start_time) IN('2016', '2017') AND status='Approved' AND Re
ason = 'Forced'
GROUP BY status, reason, year(start_time)
ORDER BY year, reason
```

Question 1.2: Building upon the query you completed in the previous question, calculate the proportion of outages that were **forced** for both 2016 and 2017. Order from **2016** to **2017**.
The output of your query should have the following column headers:
Please consider the usage of SUM(CASE WHEN Column = "Condition" THEN 1 ELSE 0 END) when writing your query.

| Total_Number_Forced_Outage_Events | Total_Number_Outage_Events | Forced_Outage_Percentage | Year |
|---|---|---|---|
| | | | |

```
SELECT SUM(CASE WHEN reason ='Forced'
          THEN 1 ELSE 0 END)
              AS Total_Number_Forced_Outage_Events,
       COUNT(*) AS Total_Number_Outage_Events,

       ROUND(100 * (SUM(CASE WHEN reason ='Forced'
          THEN 1 ELSE 0 END) / COUNT(*)), 2) AS Forced_Outage_Percent
age,

       year(start_time) AS Year
FROM AEMR
```

```
WHERE status='Approved'
GROUP BY Year
```

## 2) What was the average duration for a forced outage during both 2016 and 2017? Have we seen

Question 2.1: Write a SQL statement to calculate the AVERAGE duration of forced outage events rounded to 2 decimal places, as well as the AVERAGE amount of energy lost (MW) (also rounded to 2 decimal places) for both 2016 and 2017 due to forced outages. Don't forget to order this by YEAR.
Note:

- After the average, make sure you round to two decimal places.
- Use **TIMESTAMPDIFF** to calculate the difference between Start_Time and End_Time.

The output of your query should have the following column headers:

| Status | Reason | Year | Avg_Outage_MW_Loss | Average_Outage_Duration_Time_Minutes |
|---|---|---|---|---|

```
SELECT Status,
       year(start_time) AS Year,
       ROUND(AVG(Outage_MW), 2) AS Avg_Outage_MW_Loss,
       ROUND(AVG(ROUND((TIMESTAMPDIFF(MINUTE, Start_Time, End_Time)),2
)),2) AS Average_Outage_Duration_Time_Minutes
FROM AEMR
WHERE status ='Approved' AND reason = 'Forced'
GROUP BY year
ORDER BY year
```

Question 2.2 Write a SQL statement to compare the AVERAGE duration of each **individual outage event** (Forced, Consequential, Planned and Opportunistic) for both 2016 and 2017. Order from **2016** to **2017**.
Note:

- After the average, make sure you round to two decimal places.

- Use **TIMESTAMPDIFF** to calculate the difference between Start_Time and End_Time.

The output of your query should have the following column headers:

| Status | Reason | Year | Avg_Outage_MW_Loss | Average_Outage_Duration_Time_Minutes |
|--------|--------|------|--------------------|--------------------------------------|

```
SELECT Status, Reason,
       year(start_time) AS Year,
       ROUND(AVG(Outage_MW), 2) AS Avg_Outage_MW_Loss,
       ROUND(AVG(ROUND((TIMESTAMPDIFF(MINUTE, Start_Time, End_Time)),2
)),2) AS Average_Outage_Duration_Time_Minutes
FROM AEMR
WHERE status ='Approved'
GROUP BY status, Reason, year
ORDER BY year, Reason
```

# 3)Which energy providers tend to be the most unreliable?

Question 3.1: Write a SQL Statement to calculate the AVERAGE duration and AVERAGE energy lost (MW) for all **approved outages where the reason is equal to Forced** for each participant code, Ordered By AVERAGE energy loss (Avg_Outage_MW_Loss) and Year in descending order.

The output of your query should have the following column headers:

- Participant Code
- Status
- Year
- Avg_Outage_MW_Loss
- Average_Outage_Duration_Time_Days

**Please note that we would like you to convert the Average_Outage_Duration_Time_Minutes column to Days (*Average_Outage_Duration_Time_Days*)** (I.e. Divide the number of minutes by 60 to get hours, and then 24 to get days). Also, please ensure all your calculations are **Rounded** to two decimal places.

```
SELECT  Participant_Code , Status,
        Year(Start_Time) AS Year,
        ROUND(AVG(Outage_MW),2) AS Avg_Outage_MW_Loss,
        ROUND(AVG(TIMESTAMPDIFF(MINUTE, Start_Time, End_Time)/60)/24,2
) AS     Average_Outage_Duration_Time_Days
FROM  AEMR
WHERE Status='Approved' AND Reason='Forced'
GROUP BY  Participant_Code, Status, Reason, Year
ORDER BY Year , Avg_Outage_MW_Loss DESC
```

Question 3.2:
Write a SQL statement to calculate the Average Outage (MW) Loss and Overall
Summed Outage (MW) loss for each participant code where the Status
is **Approved** and the Outage Reason is **Forced** across both 2016 and 2017.

This data extract should be sorted in **descending order** for Summed_Energy_Lost, and
ordered from 2016 to 2017.
The output of your query should have the following column headers. Please note that,
as this is a tricky query, we've included an example of what the output would look like
for one participant code:
Please also note that for grading purposes, you should **use the YEAR() function** when
extracting the Year data from the Start_Time Column as opposed to EXTRACT().
PS: Don't forget to use ROUND() to ensure your results are rounded to 2 decimal
places!

| Participant_Code | Facility_Code | Status | Year | Avg_Outage_MW_Loss | Summed_Energy_Lost |
|---|---|---|---|---|---|
| ABC | Unit_with_highest_energy_Lost_per_participant_code | Approved | 2016 | 49.69 | 20000 |

0

Solution Query
```
SELECT Participant_Code, Facility_Code,Status, Year(Start_Time) AS Year
        ,ROUND(AVG(Outage_MW),2) AS Avg_Outage_MW_Loss
        ,ROUND(SUM(Outage_MW),2) AS Summed_Energy_Lost
FROM AEMR
WHERE Status='Approved' AND Reason='Forced'
GROUP BY Participant_Code, Facility_Code, Status, Year
ORDER BY Year ASC, Avg_Outage_MW_Loss DESC
```