

Machine Learning del proyecto

Javier Ramos Fernández

Consultas de PostgreSQL útiles

A continuación se exponen consultas útiles llevadas a cabo para realizar un procesamiento de los datos que tenemos en la base de datos *tfgdatosmodificados*, con el objetivo de crear una serie de tablas para realizar las predicciones que se nos proponen hacer en la competición.

Primera aproximación

- Extraer la hora de un valor de tipo *timestamp*:

```
SELECT EXTRACT(hour from time_window[1])  
FROM travel_time_intersection_to_tollgate_modified ;
```

En esta consulta se accede a la tabla donde se guarda el tiempo promedio de viaje en cada una de las rutas en intervalos de 20 minutos, se selecciona el primer valor del intervalo (de tipo *timestamp*) y se extrae la hora.

- Obtener el tiempo meteorológico en los intervalos de 20 minutos en los que se proporciona el tiempo promedio de viaje.

```
SELECT *  
FROM weather_data_modified w JOIN  
travel_time_intersection_to_tollgate_modified t  
ON w.date_ = t.time_window[1].date  
WHERE w.hour BETWEEN (SELECT EXTRACT(hour from  
time_window[1])) AND (SELECT EXTRACT(hour from  
time_window[2]));
```

En esta consulta se realiza una combinación de la tabla de datos meteorológicos con el tiempo promedio de viaje en intervalos de tiempo de 20 minutos, pero se seleccionan aquellas filas en las que la fecha sea igual tanto del tiempo meteorológico como de la ventana de tiempo de 20 minutos correspondiente y en la que la hora del tiempo meteorológico se encuentre entre las horas de la ventana de tiempo de 20 minutos.

- Comprobación del correcto cálculo del promedio de los tiempos medios de viaje de las dos horas previas al intervalo de tiempo de 20 minutos (comprobación de la columna *two_hours_previous*).

```
SELECT AVG(avg_travel_time) FROM  
travel_time_intersection_to_tollgate_modified  
WHERE intersection_id = 'C' AND tollgate_id = 1 AND  
(time_window[1] BETWEEN TIMESTAMP '2016-07-27 13:00:00' -  
INTERVAL '2 hours' AND TIMESTAMP '2016-07-27 13:00:00' -  
INTERVAL '20 min') AND  
(time_window[2] BETWEEN TIMESTAMP '2016-07-27 13:20:00' -  
INTERVAL '2 hours' AND TIMESTAMP '2016-07-27 13:20:00' -  
INTERVAL '20 min');
```

En esta consulta se accede a aquellas filas de la tabla *travel_time_intersection_to_tollgate_modified* en las que la ruta es C-1 y los intervalos de tiempo de 20 minutos se encuentran dentro de las 2 horas previas a la fecha **2016-07-27 13:00:00**. Una vez realizada la búsqueda de dichas filas, se procede a realizar la media de los tiempos medios de viaje de esas filas. Este valor se utiliza para comprobar en la tabla *travel_time_intersection_to_tollgate_modified*, en esa ruta y en el intervalo **2016-07-27 13:00:00' - 2016-07-27 13:20:00'** se encuentra dicho valor en la columna *two_hours_previous*.

*Nota: El cálculo de toda la columna **two_hours_previous** se encuentra en el script que se encarga de cargar todas las tablas de la base de datos **tfgdatosmodificados**.*

- Creación de una vista sobre la tabla de datos meteorológicos:

```
CREATE OR REPLACE VIEW weather_byday AS SELECT date_,  
AVG(pressure) as pressure, AVG(sea_pressure) as sea_pressure,  
AVG(wind_direction) as wind_direction, AVG(wind_speed) as  
wind_speed, AVG(temperature) as temperature, AVG(rel_humidity) as  
rel_humidity, AVG(precipitation) as precipitation  
FROM weather_data_modified GROUP by date_  
ORDER BY date_;
```

En esta consulta, para cada uno de los días de entrenamiento proporcionados por la competición, se calcula el promedio de los valores de cada uno de los

atributos meteorológicos, de tal forma que esta vista nos sirve para poder tener a mano datos meteorológicos por día y con el objetivo de poder unirla con otra tabla.

En esta primera aproximación no se ha tenido en cuenta si el día es festivo o laboral puesto que no se dispone de dicha información.