Minería de datos con twitter y mensajes para sensores MQTT

Dulce María González Reyes A01745835

Fernando Joshua Alvarado Ortiz A01748693

Jorge Isidro Taymani Gates Zuckerberg A01745907

José Luis Madrigal Sánchez A01745419

Alejandro Bastida Cortés A01746457

José Ángel García Gómez A01745865

Mayo, 2020

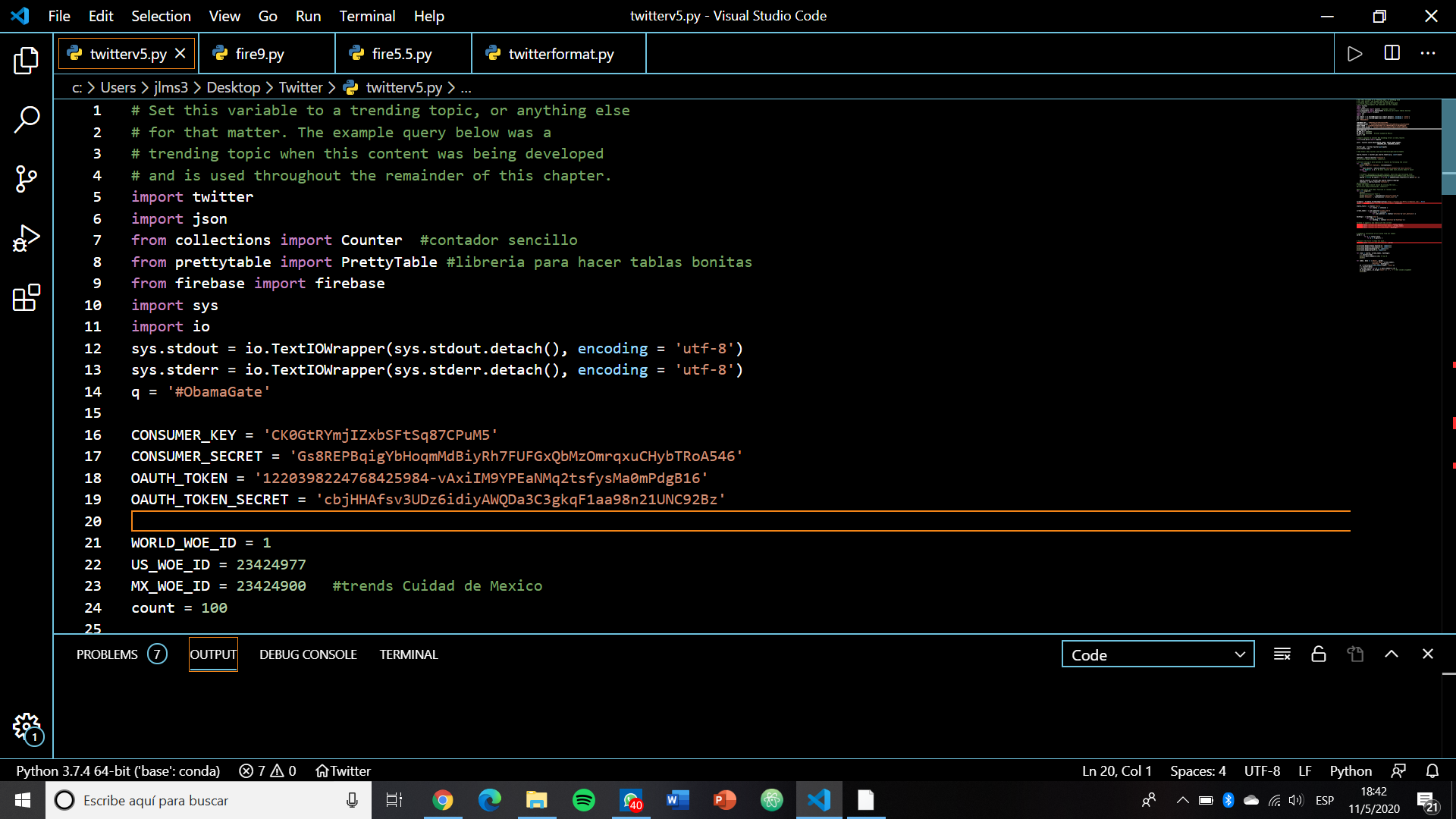
Instituto Tecnológico y de Estudios Superiores de Monterrey

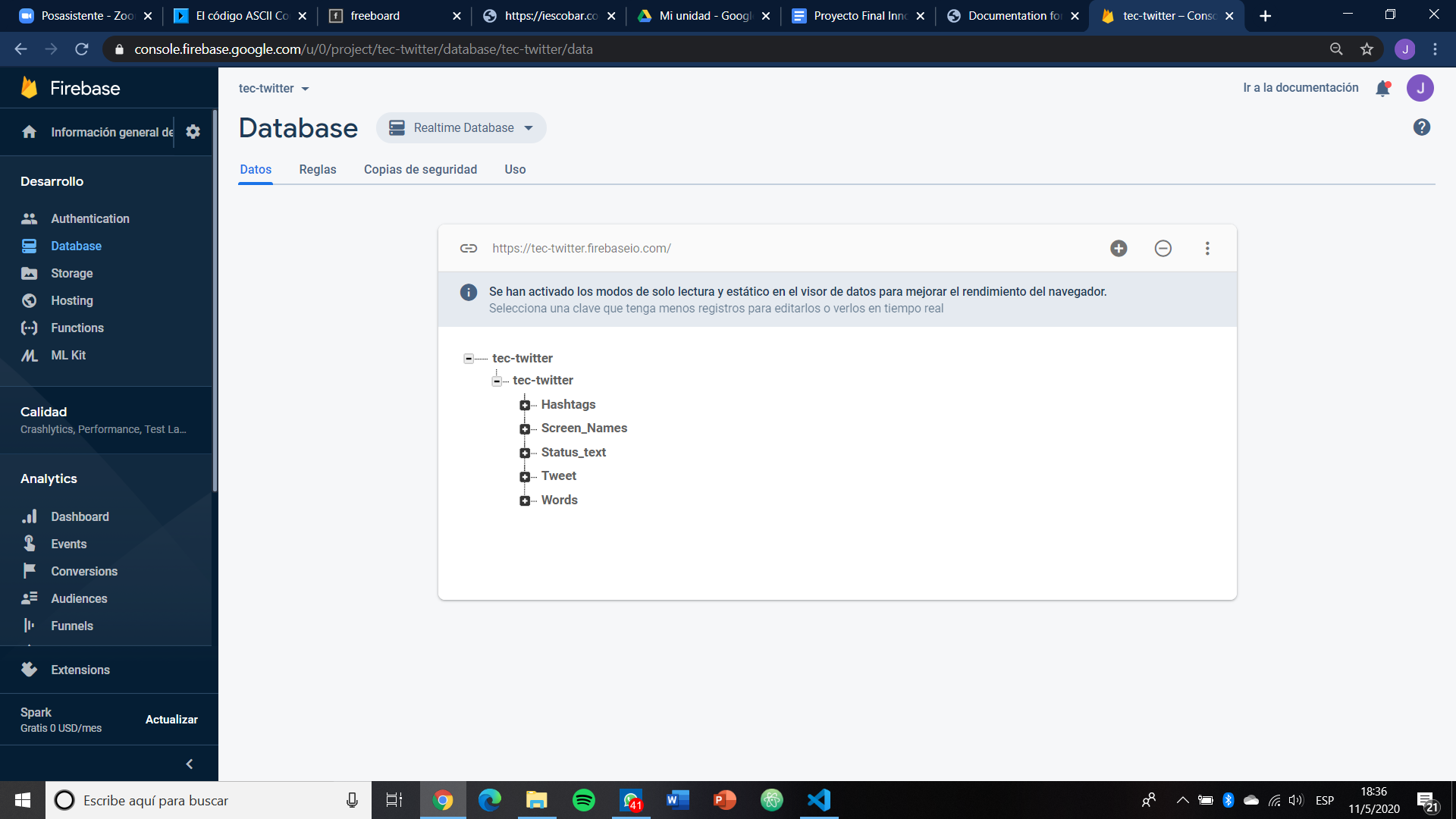
Innovación tecnológica

Departamento de Ciencias

**¿Qué hicimos?**

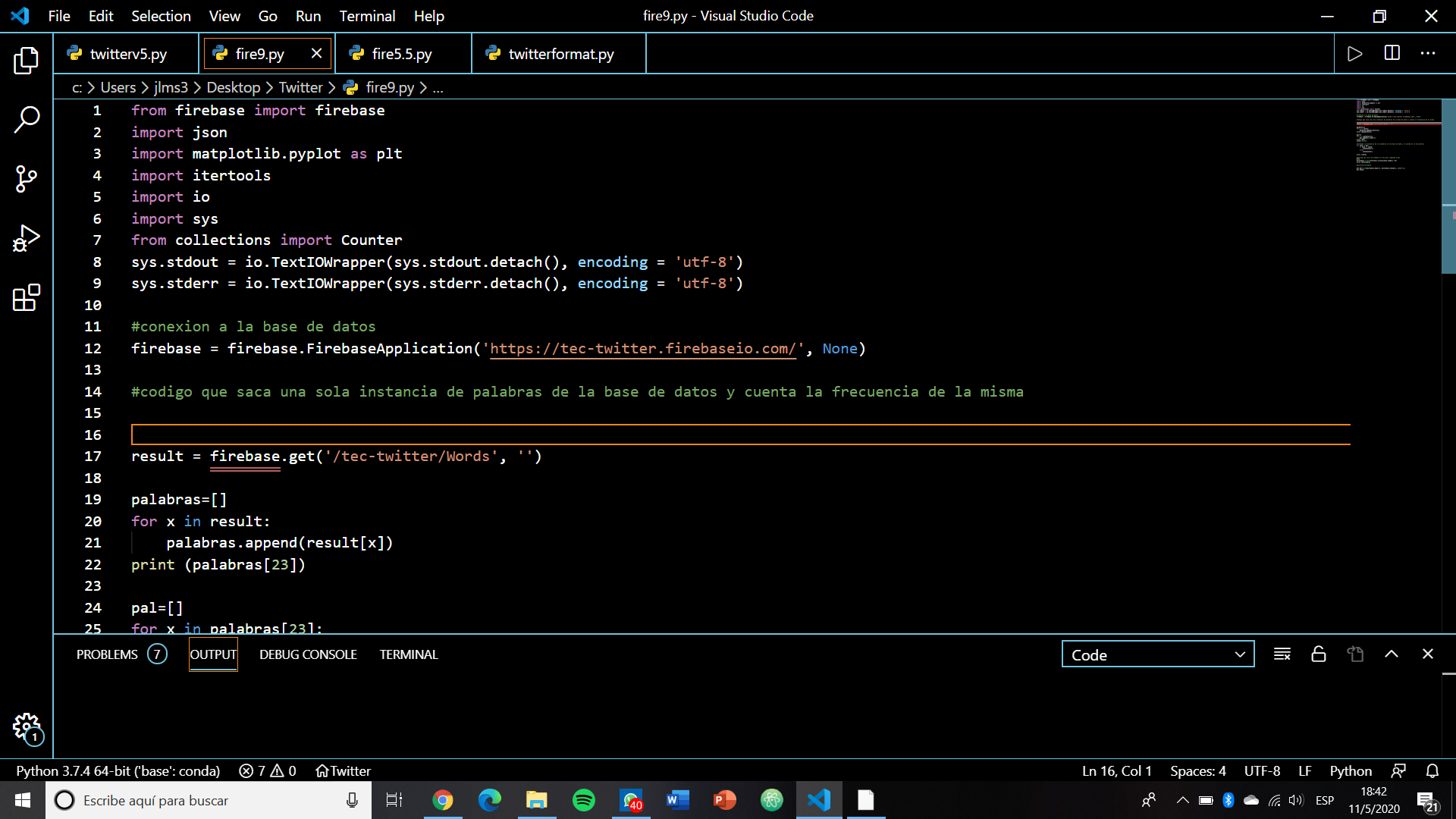
Primero que nada, creamos una base de datos con firebase en la cual podríamos suministrar información. Luego por medio de un código de Python y Twitter API, fuimos colocando en la base diversos trending topics durante varios días. Básicamente se sacaba toda la información de los hashtags, como las palabras, el screen name, tweets, etc. A continuación se muestra el código utilizado y la firebase con los datos:

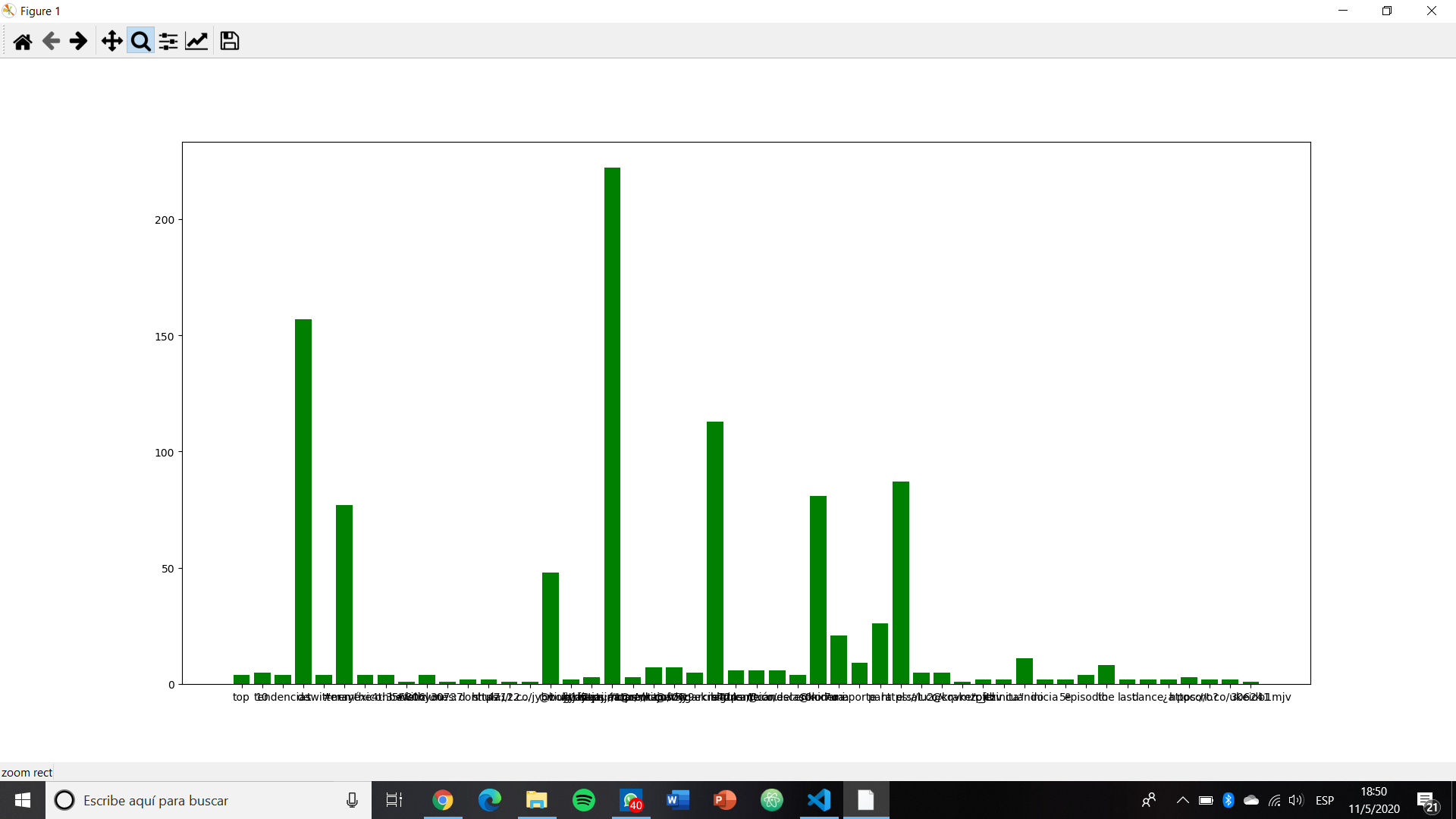




**Resultados**

Después de tener los hashtags, hicimos un conteo de las palabras más repetidas en todos ellos, mostrando las cifras por medio de un histograma. Asimismo, nosotros podíamos establecer cuál número de palabras queríamos visualizar, en nuestro caso, terminamos poniendo 50, ya que queríamos tener un panorama más amplio para poder encontrar cosas más interesantes. Por lo tanto, se muestra el código y el histograma obtenido:

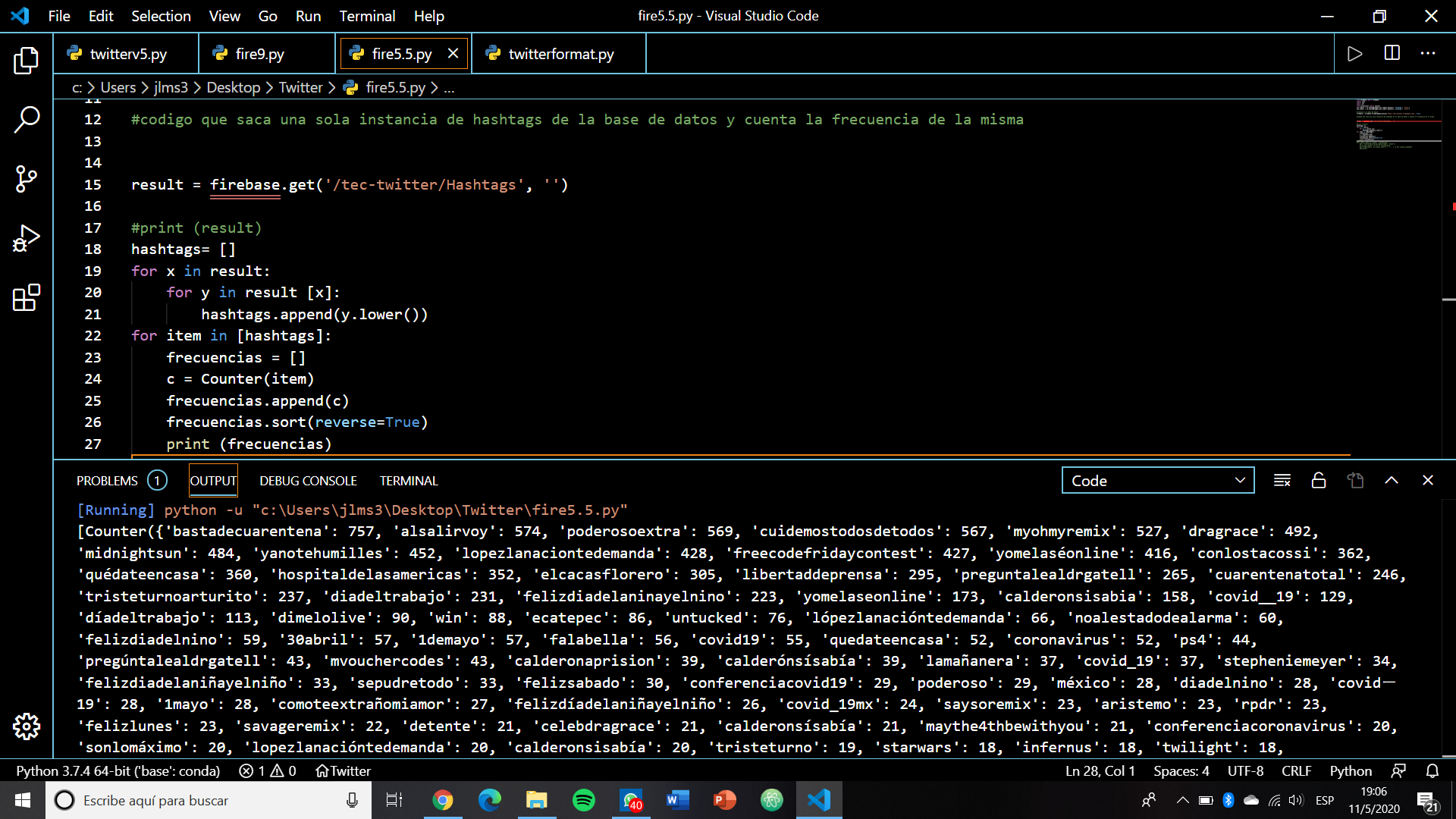




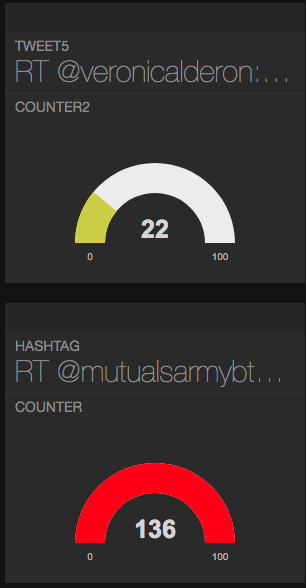
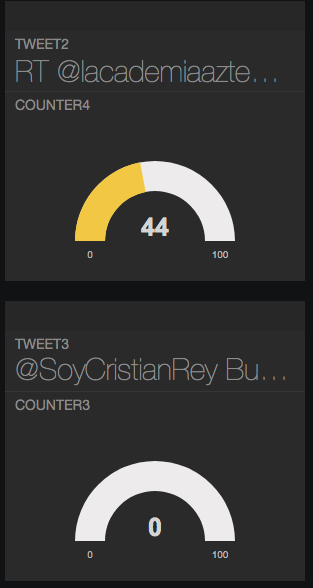
Ya que no se pueden visualizar muy bien las palabras, se explicarán de manera general. La que obtuvo una mayor frecuencia, siendo mayor a 200, fue la de #tristeturnoarturito que fue tendencia desde el 4 de mayo por el Star Wars Day. En cambio, las demás palabras con muchas repeticiones son básicamente artículos y preposiciones del español, como “en”, “la”, “de”. Aunque también se puede ver la palabra “twitter” y el “rt por obvias razones, e inclusive varias palabras forman un pequeño fragmento de algún mismo tweet, como (“mi”, “aporte”, “para”, “el”) o (“piel”, “chinita”, “cuando”, “inicia”, “5°”, “episodio”) o (“the”, “last”, “dance”), siendo el último ejemplo una relación con la serie de los Bulls de la NBA en Netflix.

**Información recaudada**

Por último, por medio de otro código de python, generamos un contador de los hashtags de la base y mostramos las frecuencias de manera descendente, para que así pudiéramos ver aquellos que se repetían más. Y observamos que muchos de ellos están relacionados con la cuarentena, el COVID 19 e inclusive el gobierno. Y hasta abajo de la imágen se observa claramente que hashtags de Star Wars y tristeturno fueron tendencia también, por eso se pudieron ver muchas palabras relacionadas en las mayores frecuencias del histograma.



Después de esto, por medio de freeboard.io hicimos un dashboard con la información de la base. Cabe mencionar que tuvimos que exportar un archivo json de la firebase para que así el profesor pudiera ponerlo en su página, de la cual tomaríamos los datos. Por lo que, al estar conectados al json, pudimos hacer paneles de tipo texto con algunos hashtags o tweets, para luego poner otro de tipo gauge con el contador, para poder ver su presencia y saber su frecuencia. A continuación se muestran algunos:



**Conclusión**

Podemos decir que este proyecto nos gustó mucho, ya que pudimos conocer más sobre el área de analítica y encontrar más funciones de Python. Asimismo, fue bastante interesante el poder utilizar códigos que hicieran uso de información de una red social, ya que de verdad se puede sacar mucha información. Pero más importante, aprendimos a analizar datos y sacar cosas importantes o en todo caso, encontrar algo específico.