**Pendientes primer parcial**

**Twitter**

**César**

* Producir varios datasets en los cuales se filtren por tiempo:, weekday, weekend, mañana y noche.
  + Para con esto ver si cambia las asociaciones de las palabras en diferentes franjas horarias
* Estadística descriptivas: contar cuantos usuarios hay en el dataset, etc.
* Considerar Quito y Ecuador como stopwords.
* Sería bueno ver la temporalidad de las palabras, por ejemplo #birthdaygirl para saber si fueron en período de tiempo corto o largo, con lo cuál las palabras con temporalidad corta no interesan tanto sino las palabras con temporalidad alta.
* Con respecto a los POIs seleccionarlos a los top usando varias métricas, por ejemplo los lugares con mayor cantidad de tweets.

**Madelyne**

* Sacar el **degree promedio (weighted degree)** al grafo (grafo por cada POI en el top 5), con lo que podemos decir estos POIs con # de tweets y estos grafos con estos degreee promedio lo escogimos para categorizarlos.
  + Lo caracterizan haciendo las recomendaciones temporales. Determinar si es que hay diferencias como los usuarios hablan del POI en la mañana, noche, de weekdays y weekends
  + Poner peso a los arcos, calculando a el (weighted degree): por ejemplo hay unos que le llegan 10 pero esos 10 llegan solo una vez.
  + Identificar los nodos más importantes.
  + Hacer grafos por temporalidad
  + Identificar de algún modo cuantitativo que palabra es más relevante por bandas temporales, se puede sacar el PageRank. Es decir obtener la palabra con mayor PageRank en las diferentes bandas temporales.
* Intentar ver cuales son los usarios que tienen más tweets geolocalizados (Segundo Parcial)
* Pedir nombres de usuarios para recoger timeline y ver de esos usuarios cuanto twitean en quito (SEGUNDO PARCIAL)

**TripAdvisor**

**Kevin, José lo suyo se parece bastante si quieren ponganse de acuerdo que hace cada uno**

* Con respecto a los POIs seleccionarlos a los top usando una métrica (buscar alguna métrica que un paper coge a los top con más reviews o algo similar).
* Estadística descriptiva: obtener el numero de review en POI, promedio y otras
* Sacar de los POIs cual es el lugar con mayor diversidad de turistas locales(Quito), globales(Extranjeros o de otra ciudad de Ecuador, facil el filtro es que si en el perfil no dice Quito). No solo se puede decir que tiene más reviews sino que se puedo decir que la gente que hace es reviews es de quito o es de otro lugar de experiencia..
  + Supongamos que la gente quiere buscar una panadería que visiten solo los quiteños, sino que tambien pueden responder si es que hay una panadería que la visiten de afuera.
* Graficar de todos los lugares mezclados sin categoría, los que son más locales o globales. Hacer un heatmap para los locales y uno para los globlales
* Con la información de los usuarios. Ver cuáles son los sitios que han obtenido comentario con usuarios más influyentes. Por ejemplo en Quito el lugar que recibio más comentarios locales fue esta \*\*\* mientras que el recibio más comentarios de usuarios más influyentes donde influyente son usuarios con helpful votes.
* Elegir los top lugares con mayor comentaristas con helpful votes son estos \*\*\*\*\*\*
  + De pronto para los top lugares seleccionados hacer una serie de tiempo.

Luego del scrapy, no olviden cada valor que obtengan que vaya hacer utilizado para hacer estadítica descriptiva o para graficar deben agregarlo al dataset del CSV suyo de trip advisor usando pandas.

**Gráficas**

Para cada cosa pendiente que queda piensen, en que gráfica se vería mejor para mostrar los resultados y así ponernos de acuerdo de que gráfica hace cada uno. Recuerden para las gráficas se usarán solo los CSVs.