**ROADMAP**

TRABAJO FIN DE MÁSTER

Sandra de la Fuente Cáceres

# Introducción

# Tecnologías aplicadas

# Descripción de la estructura de directorios

# Descripción de los datos de entrada

# Metodología

# Resumen del resultado

# Visualización Dashboard

# Introducción

El propósito de este Trabajo Fin de Máster (en adelante TFM) ha sido el de explorar y mostrar las posibilidades que puede ofrecer el análisis de datos, a la hora de tomar decisiones de negocio en el sector turístico español, tan importante para la economía de este país.

Los estudios existentes hasta la fecha sobre Turismo en España, se basan en su gran mayoría en encuestas o entrevistas, por ese motivo para este TFM, los datos utilizados son del Instituto Nacional de Estadística, disponibles en los siguientes enlaces:

* Estadística de movimientos turísticos en frontera. Frontur <http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176996&menu=resultados&secc=1254736195382&idp=1254735576863>
* Encuesta de gasto turístico. Egatur <http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177002&menu=resultados&secc=1254736195390&idp=1254735576863>

Los datos recopilados son de los últimos 19 meses, periodo entre Octubre del 2015 y Abril del 2017.

Objetivos del proyecto:

1. Análisis descriptivo y predictivo de la actividad turística en España, para predecir el número de turistas por nacionalidades que decidan viajar a un lugar determinado.
2. Segmentación de los turistas según características similares, para poder definir acciones de marketing específicas para cada tipo de turista, acciones publicitarias, así como lanzamiento de promociones especiales dirigidas a un turista objetivo.
3. Conocimiento holístico del turista.

Los estudios existentes hasta la fecha sobre Turismo en España se basan, en su gran mayoría, en encuestas o entrevistas a expertos promovidos por el Ministerio de Industria, Energía y Turismo como principal organización pública o bien por empresas privadas. Esto significa que, en general, el sector no dispone de datos reales de los turistas y sólo pueden extraer muestras de toda la población.

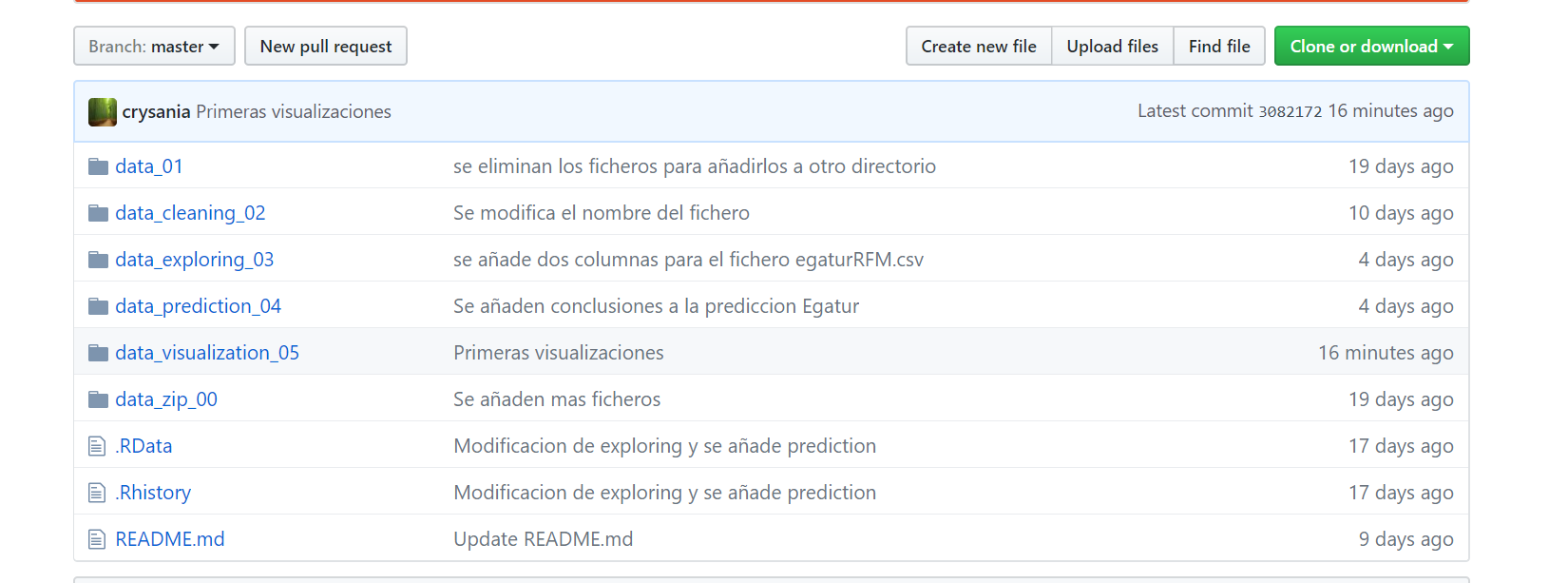
# Tecnologías aplicadas

Para este TFM se utilizan las siguientes tecnologías:

* Python (*[script\_datos\_py](https://github.com/crysania/TFM_DataScience/tree/master/data_zip_00/script_datos_py" \o "script_datos_py)*) para descomprimir ficheros de los datos obtenidos del INE.
* R para el proceso Data Engineering:
  + Limpieza de los datos
  + Análisis descriptivo de los datos
  + Análisis predictivo de los datos
* Tableau para realizar dashboards.

# Descripción de la estructura de directorios

A continuación, se describe la estructura de directorios del repo Git con el orden de ejecución de los mismos.

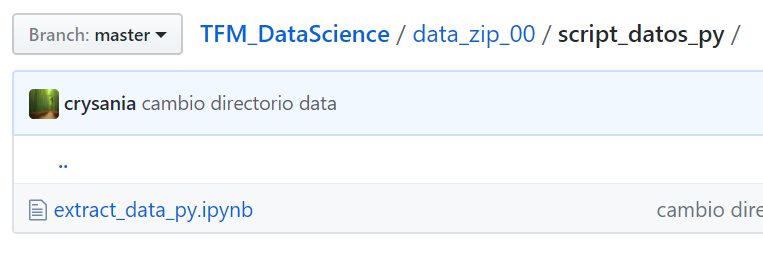


[***data\_zip\_00***](https://github.com/crysania/TFM_DataScience/tree/master/data_zip_00)*🡪* Ficheros zips del INE dentro de cada directorio EGATUR y FRONTUR.

Dentro está el directorio (*[script\_datos\_py](https://github.com/crysania/TFM_DataScience/tree/master/data_zip_00/script_datos_py" \o "script_datos_py)*) para descomprimir ficheros de los datos obtenidos del INE.



[***extract\_data\_py.ipynb***](https://github.com/crysania/TFM_DataScience/blob/master/data_zip_00/script_datos_py/extract_data_py.ipynb) 🡪 Script de Python para descomprimir los ficheros de los datos obtenidos del INE.



# Resumen del resultado

El TFM ha permitido obtener información sobre los siguientes aspectos:

* Origen principal de los visitantes.
* Origen de los turistas que optan por cada una de las Comunidades Autónomas.
* Zonas donde prefieren alojarse los visitantes extranjeros.
* Gasto medio diario y gasto acumulado a lo largo de toda la estancia.

A partir de las conclusiones que ofrece el análisis de datos realizado en este TFM, se concluye el estudio con una serie de recomendaciones tácticas y estratégicas dirigidas a los gestores hoteleros. Estas recomendaciones se enfocan a:

* Aumentar la captación de clientes y determinar en qué países es recomendable focalizar la acción comercial.
* Determinar las Comunidades Autónomas donde se realizan las transacciones comerciales.
* Garantizar un producto atractivo y adaptado a las verdaderas necesidades de los clientes (duración óptima de los paquetes de estancias, información relativa a la oferta complementaria demandada según nacionalidades...).
* Dado el volumen de turistas que aportan, se considera necesario realizar acciones de captación en los siguientes países: Francia, Reino Unido y Alemania.
* Ubicación/expansión: áreas de interés en función de las nacionalidades.
* Área tecnológico: con campañas SEM / SEO, idiomas en los que la web debería estar traducida y presencia en webs o intermediarios específicos.

# Visualización Dashboard

Para realizar el análisis de los datos y visualizar los resultados, se realiza un dashboard interactivo, la herramienta seleccionada es Tableau (debe estar instalado previamente).

Los datos a visualizar, parten de la exportación de los mismos en ficheros con formato csv.

## EGATUR

El dashboard generado se obtiene del fichero: " RFM\_EGATUR\_CLUSTERS\_6.csv".