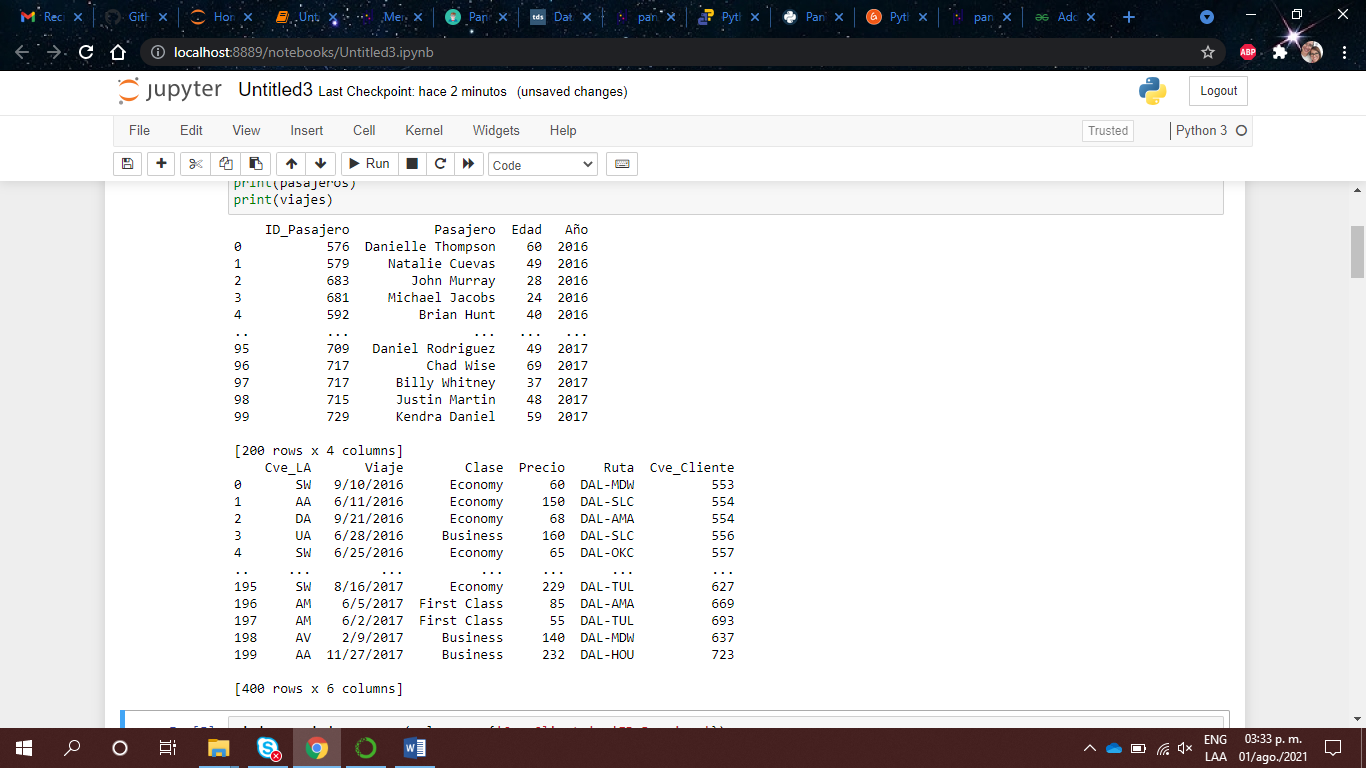
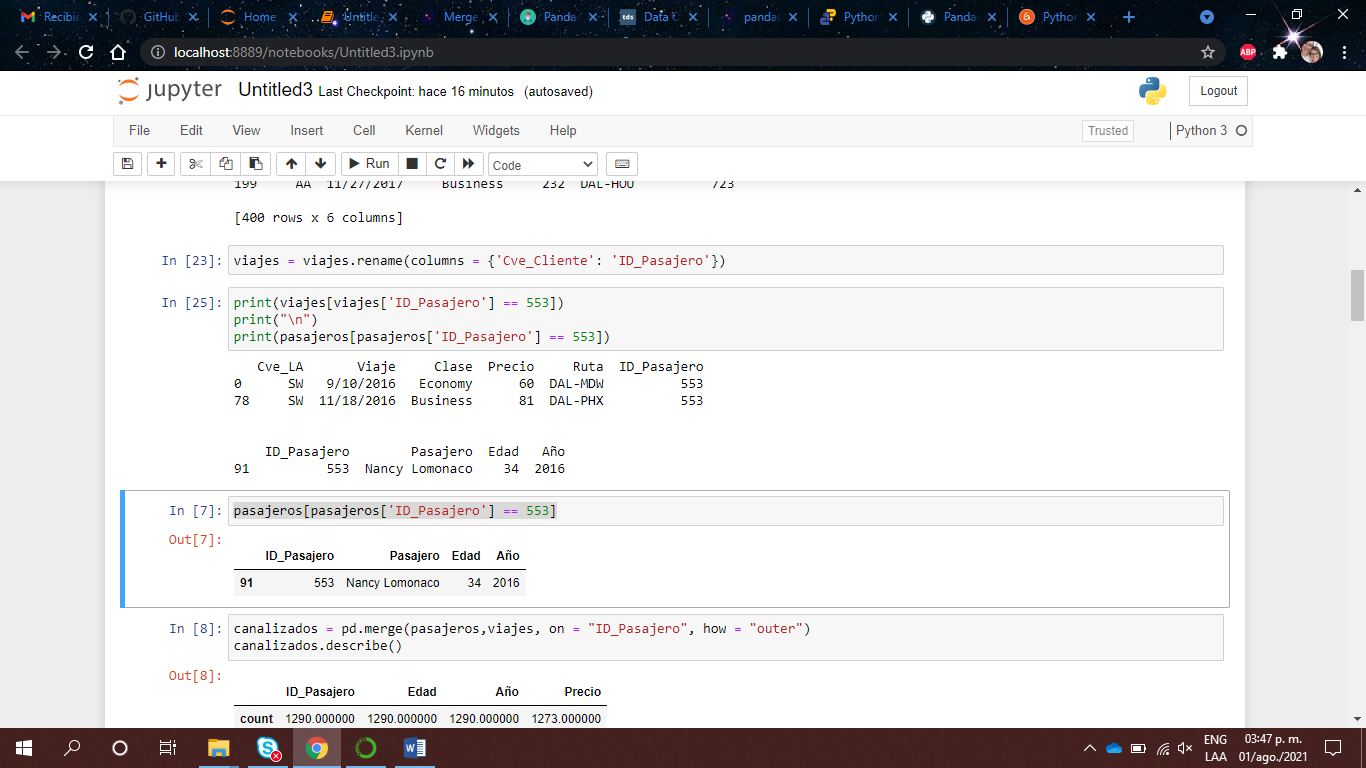
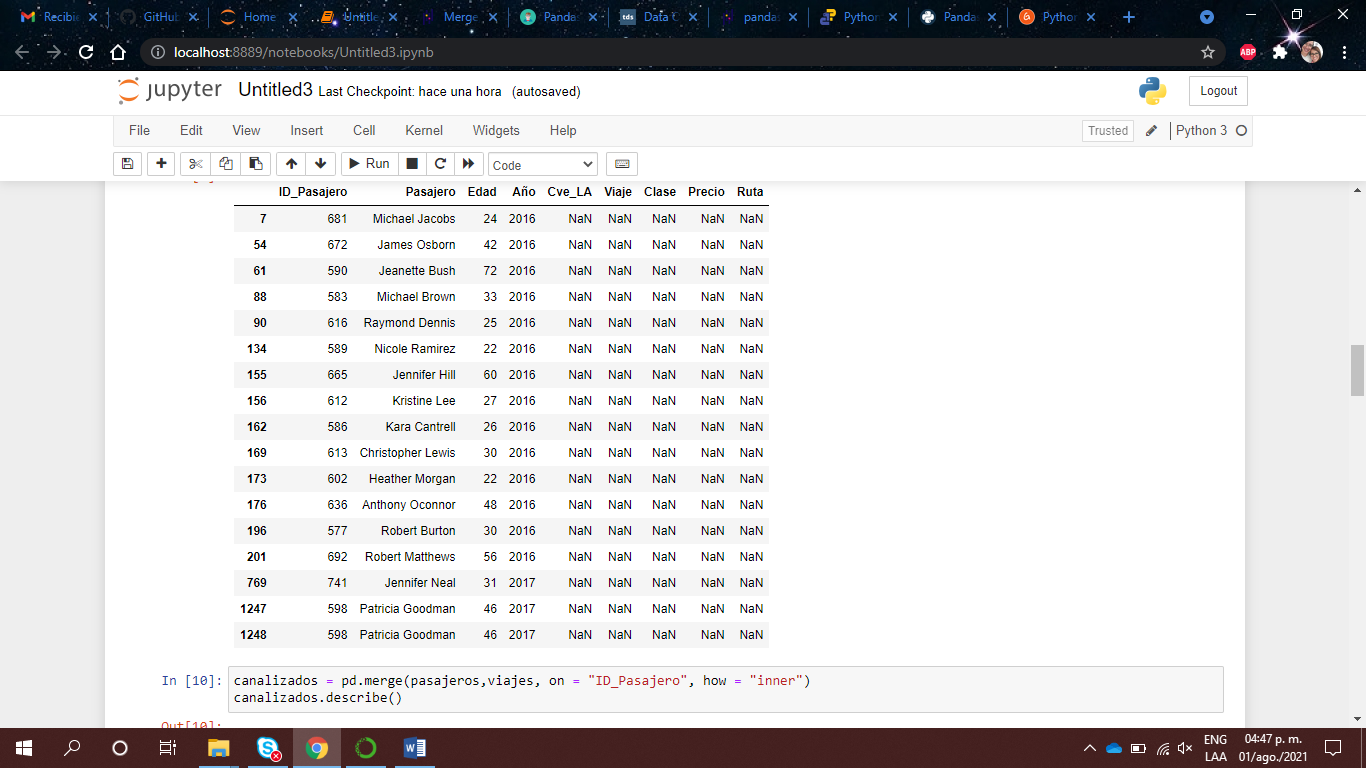
1. Lo primero que se hizo fue obtener los datos de los link que se me proporcionaron. Identifiqué que los datos tenían un formato parecido a datos tipo json, por lo que se le dio un formato de archivo json, de manera que fuera posible leerlos con facilidad.
2. Una vez los datos fueron leídos se procedió a juntar los conjuntos de datos, los pasajeros y los vuelos. Para juntar a los vuelos, se observó que los dos conjuntos de datos tenían la misma estructura, por lo que se utilizó la función “concat” para juntar ambos conjuntos de datos. Para los pasaron, se observó que, por defecto, los datos no venían con un campo de año, que los pudiera identificar entre sí, por lo que antes de juntar los conjuntos de datos, se le agregó a cada tabla de pasajeros un campo con el año, para que se pudieran identificar.



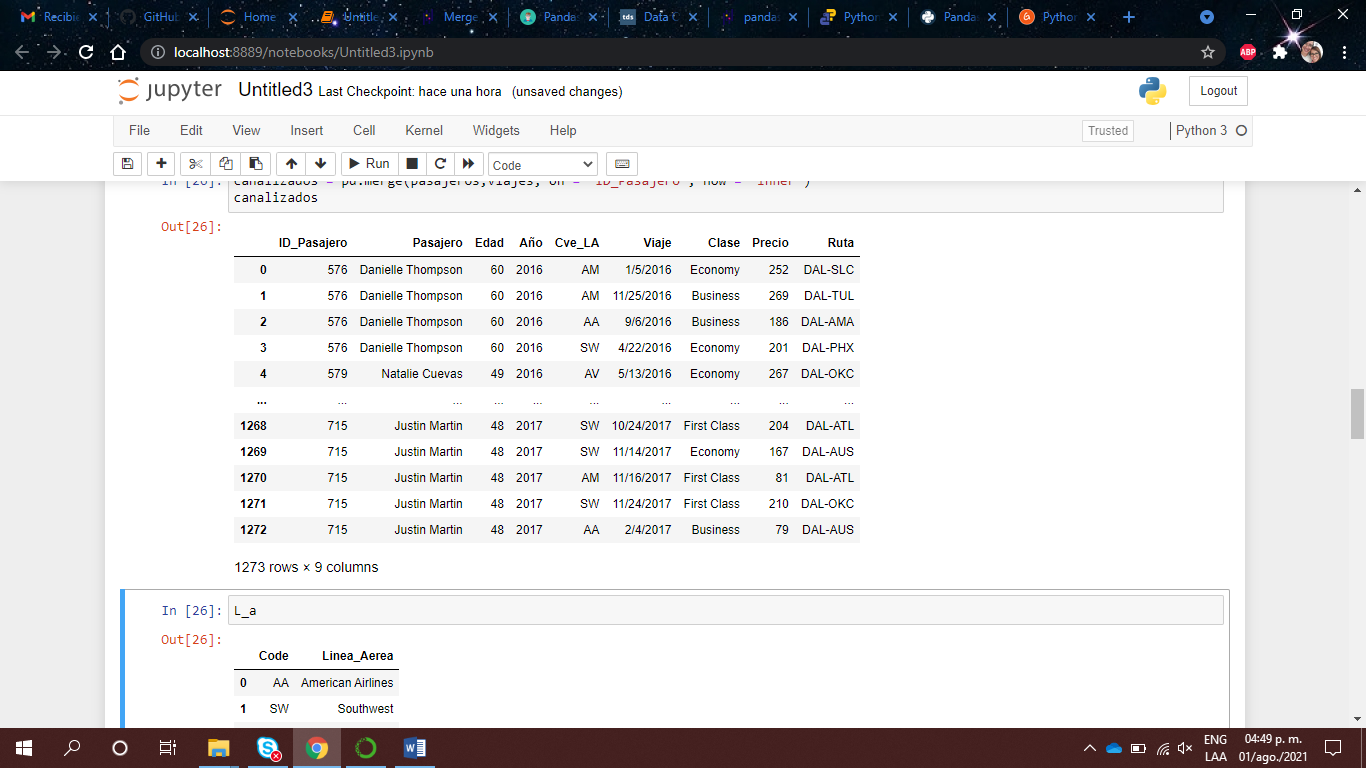
1. Antes de intentar juntar las tablas de pasajeros y vuelos, se tuvo que ver la cardinalidad que tenían los datos entre si, para eso primero se identificó el campo con el que se relacionaban ambas tablas, en este caso fue “ID\_Pasajero” “Cve\_Cliente”, después busqué un código de cliente en particular para ambas tablas, y se observó que, en la tabla de pasajeros, solo hay un pasajero por código, pero hay varios vuelos por código, por lo que relación entre ellos es de “uno a varios”.



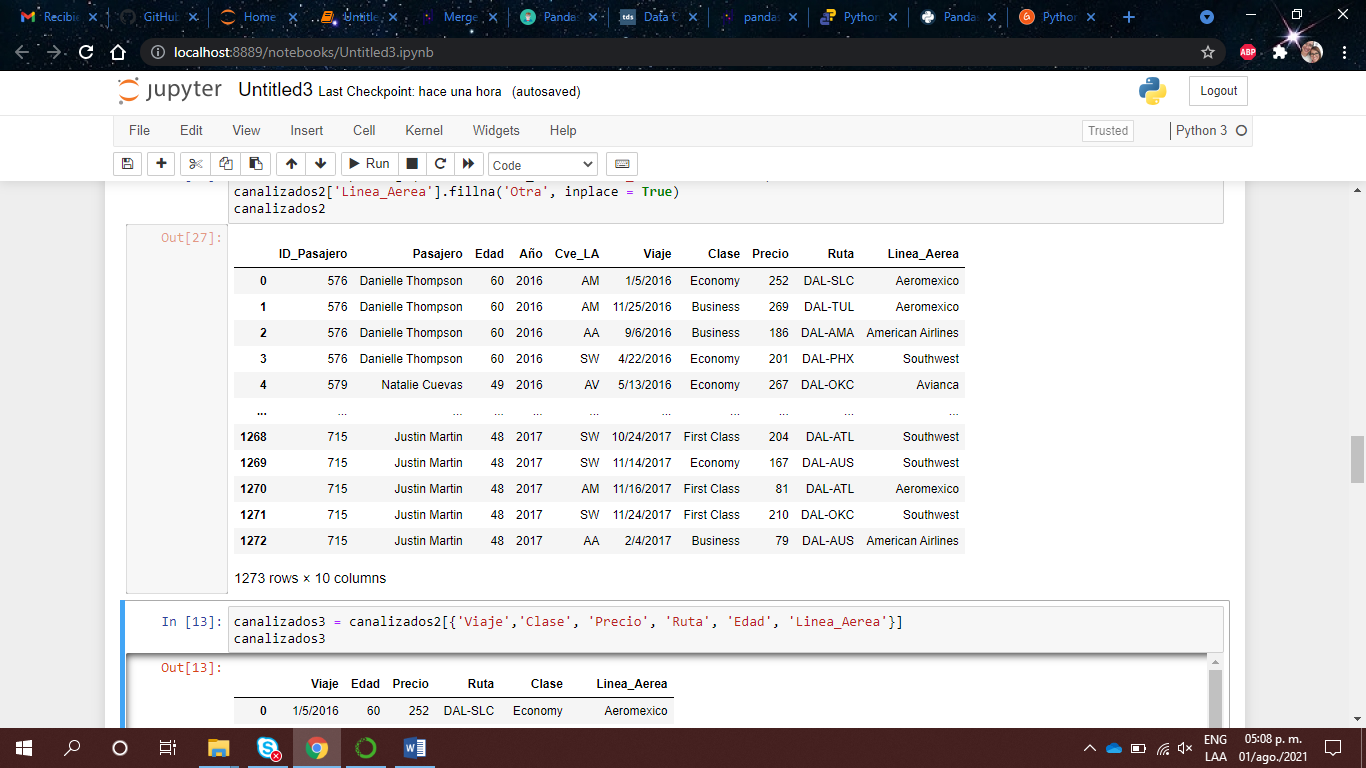
Para juntar ambas tablas se utilizó la función “merge”. Se encontró que había algunos pasajeros que no se encontraban en la tabla de vuelos.



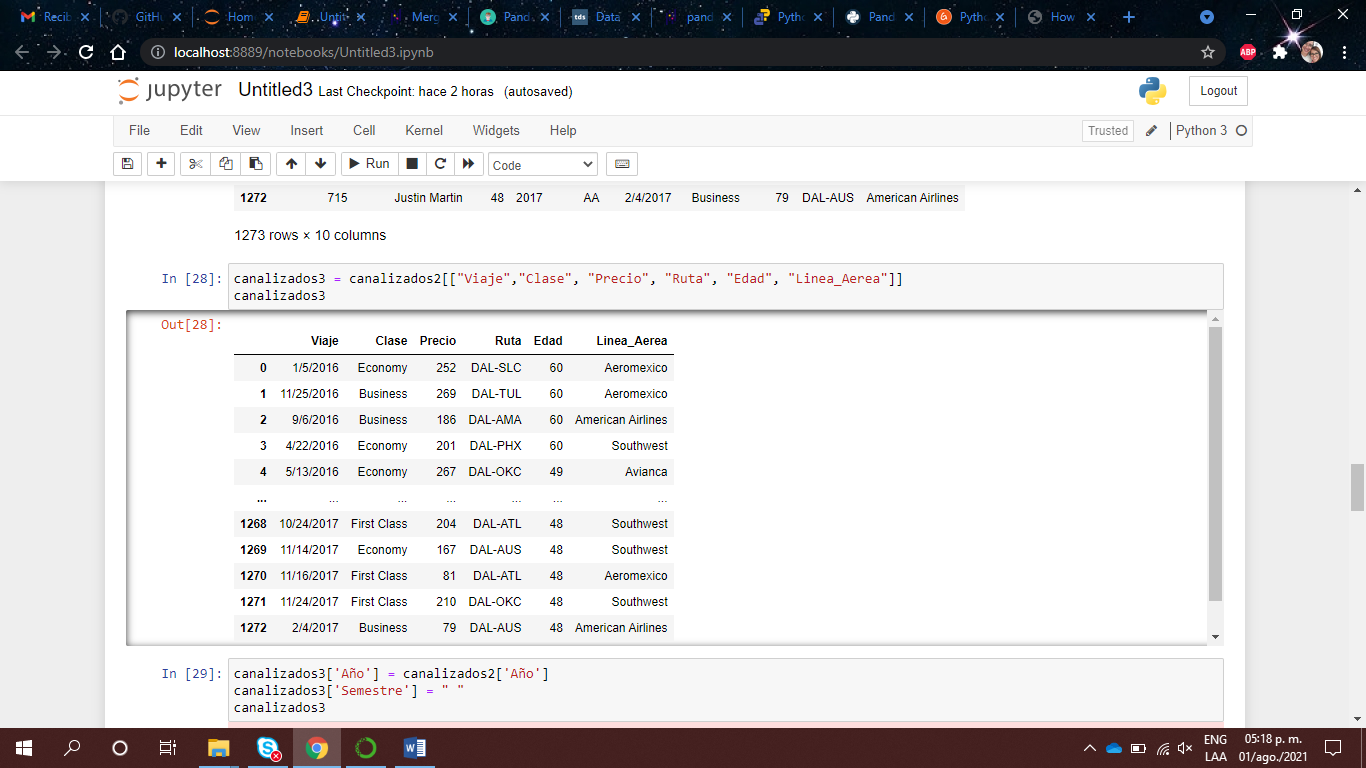
Debido a lo anterior, al momento de consolidar la información de ambas tablas, se hizo un “inner merge” para que en la tabla consolidada no hubiera registros de clientes sin información de vuelo.



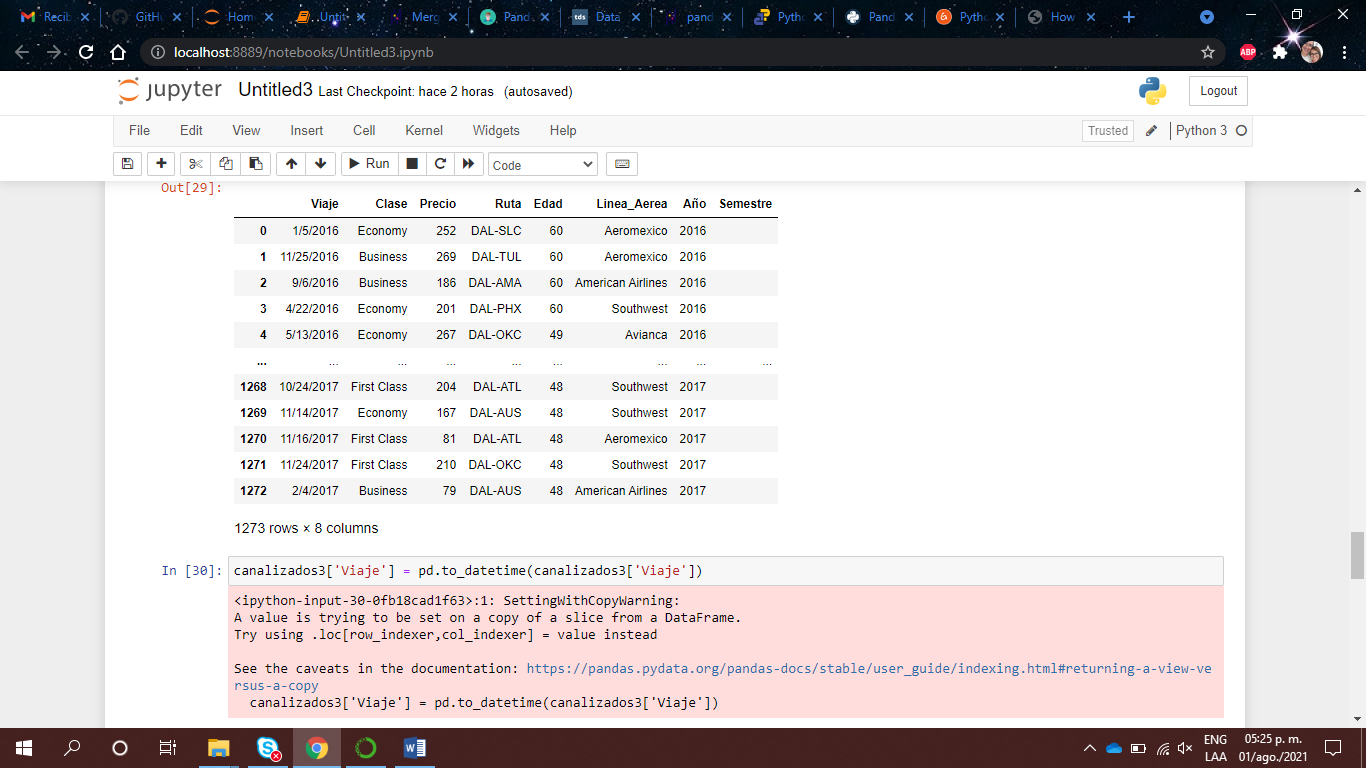
1. Para poder juntar la información de las líneas aéreas con la tabla consolidada de pasajeros y vuelos, se identificó con que campo se pueden relacionar ambas tablas, en este caso fue la columna “Code” de la tabla de líneas aéreas y la columna “Cve\_LA” de la tabla de los datos consolidados. Por lo que se hizo un “left merge” con la tabla de los datos consolidados y la tabla de líneas aéreas, de manera que quedara una tabla con la columna agregada de línea aérea, y aquellos campos que no tenían línea aérea especificada, se actualizaron esos datos de manera que la línea aérea fuera “otra”.



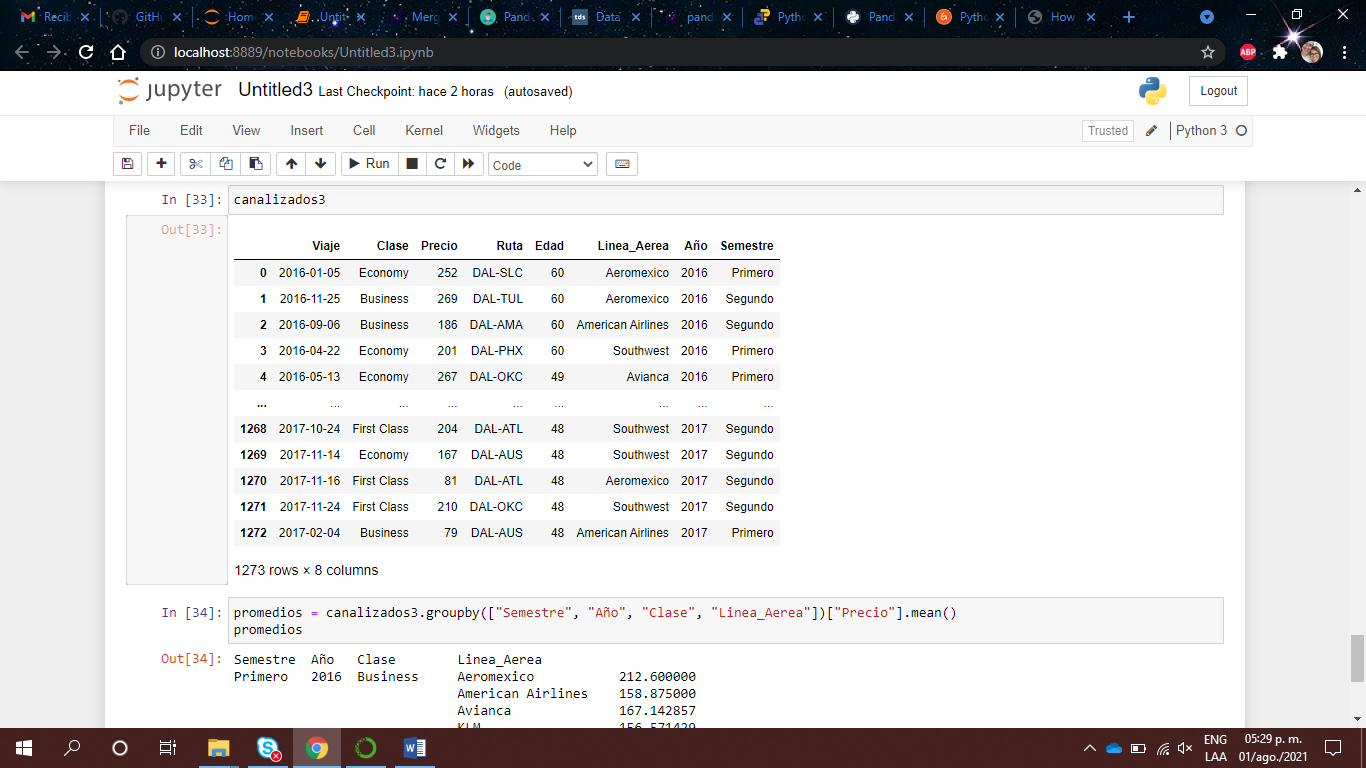
Después de tener la tabla con el agregado de la línea aérea, a partir de ella se creó otra tabla en donde se seleccionaron solo los valores que se me pidieron: Fecha del viaje, Clase, Precio, Ruta, edad y Línea aérea.



1. Para poder obtener los promedios agrupados, lo que se hizo primero fue agregar dos columnas a la tabla consolidada anteriormente creada. Una columna del año, que se obtuvo tomando la columna año de una de las tablas consolidadas que se tenían, y otra columna de semestre la cual provisionalmente se llenó con un vacío.



Después de eso se cambió el formato de la fecha del viaje por un “datetime” esto se hizo para facilitar las operaciones para identificar a que semestre pertenece cada dato. La columna “Semestre” se terminó llenando con “Primero” cuando pertenece al periodo del primer semestre del año, y “Segundo” cuando pertenece al periodo del segundo semestre del año.



Ya con la tabla con el formato que se buscaba, se pudo realizar el promedio de precios agrupados por Semestre, Año, Clase, Ruta y las Línea Aérea.

