Given the Policy and Premium tables in separate sheets within the same excel file.

Summary of data exploration:

* 3 months’ worth of transactions
* 2 types of LOB: Travel and PA
* ENDO and CANC are the only negative GWP
* Information utilised will be:  
      - Policy table:  
          - CHDRNUM (can be from either table)  
          - LOB  
      -Premium Table:  
          - TRANTYPE  
          - GWP  
          - YrM

Deductions and Assumptions:

* Both Policy and Premium tables are indexed to be exactly matched
* All the data is correct and there are no nulls
* Year and Month will be extracted from the transaction date from the YrM
* D\_tran, D\_eff, D\_com, D\_exp are unnecessary for this analysis
* Since definition of GWP Cancelled and Booked are not clear, they will be assumed:
  + GWP Cancelled will be a SUM of the GWP where TRANTYPE = CANC sorted by LOB and by Month
  + GWP Booked will be a SUM of the GWP where TRANTYPE != CANC sorted by LOB and by Month
* As such, the assumption of all TRANTYPE available in the dataset are the only TRANTYPE possible.

Methodology:

1. Excel file read and separated into dataframes
2. Necessary data extracted and combined to a single dataframe (***data***)
3. Based on the assumption that transaction type booked is everything that is not cancelled, all non-cancelled transaction type has been renamed to booked
4. Filter ***data*** by start of year to current month and start of previous year to the same month.
5. Create an empty dataframe(***final\_df***) with length based on the filtered ***data*** above
6. Create a function to aggregate the GWP values of cancelled and booked in accordance to Line of Business, month and year of filtered ***data***  and fill in the empty ***final\_df*** accordingly
7. Apply function to filtered ***data*** and fill in empty ***final\_df***
8. Export ***final\_df***  as a csv file
9. Using APScheduler and Flask, automate the running of the above report generator by last day of the month.

Note:

Code was written to be robust on the assumption that there are only 2 tables in separate sheets within the single excel file being POLICY and PREMIUM.

It matters not the length or variety of Line of Business, however the 2 sheets have to be matched in index.

The virtual environment for the flask app and APScheduler to run will have to be always active in this scenario.