1. **You have some files in the ‘data’ directory. The data in these files looks like this:**
    a. “20-Aug-2020,Tax,‘8,900’,‘1,000.00’,‘90,000’”
    b. The last three columns are ‘Deposits,’ ‘Withdrawals,’ and ‘Balance’
    c. **Write a Spark structured streaming program** that will read the files as a file stream (one file at a time)
    d. Convert the string values into float values
    e. Send the stream to console output mode

Hints: Use file stream as input on the 'data’ directory
Read df and select fields after providing a schema
Use write console mode for write stream

2. **Continuing with the problem above, now you have converted string values to float.**
    a. Write a Spark streaming program to sum the Deposits (column in the file) on each Description (column in the file) type

Hints: Use file stream as input on the ‘data’ directory
Read df and select fields after providing a schema
Apply a group by operation on the Description column and use the agg method sum
Use write console mode for write stream

3. **Continuing with the first problem, now you have converted string values to float.**
    a. Write a Spark streaming program to filter all the data under Description (column in the file) that belong on ‘Tax’
    b. Average out the ‘Balance’ field and rename it as average_balance
    c. Display the output in the console write stream

Hints: Use file stream as input on the ‘data’ directory
Read df and select fields after providing a schema
Apply a filter operation on the Description column with a value and use the agg method avg
Use write console mode for write stream

4. **Continuing with the first problem, now you have converted string values to float**
    a. Consider creating a static dataframe using the JSON data given below:

```
'[{"type":"Tax","t_type":"offline"},{"type":"Cash","t_type":"offline"},{"type":
"NEFT","t_type":"online"}]'
```

    b. Write a Spark streaming program to add a new column to the streaming data ‘t_type’ using the static data frame given above
    c. Use a join between the streaming df and the static df

Hints: Use file stream as input on the data directory
Read df and select fields after providing a schema
Create a new static df using JSON and join both the dfs as a left join
Use write console mode for write stream

5. **Here, we want to create a window batch of time event 5 mins.**
    a. Create a Rate streaming that will generate 1 row per second
    b. Extract the value and the timestamp of the row generated
    c. Write a Spark structured streaming program that will generate a count of values generated per minute and write the output every 30 seconds
    d. Display the output in the console write stream

Hints: Use rate stream as input, which will generate a dummy row every second
Read df and extract fields such as timestamp and value
Apply groupby on a window of 1 min and use the count method for obtaining the count
Use write console mode for write stream