



Informatica

**SET - C**

**Name :- Talebul Islam By**

**Reg No :- 11904195 Amanpreet Kaur Ma'am**

**CA :- 03**

**File link: - <https://github.com/talebulislam/SSIS>**

Q1) Using TITANIC DATASET(available in SAMPLES folder of rapid miner repository) perform the following questions

Question no:- 1.1:- Create a process to remove the missing values and delete the column "cabin" and "Part of Embarkment".

Ans:- Step 1:- Open the Rapid miner software.

→ Open the Turbo prep and load the Titanic dataset.

\* first I removed the missing values and replace the value in age column 0.99

\* In the column life boat removed the missing value and I put "null"

\* I delete the cabin and Part of Embarkment column

\* Passenger fare column removed the missing values and I put 0.235

\* In the last step I create the process and export the process.

DOMS and question-1  
Answers.

## 1. Remove Missing values from life Age and I putt value “0.99”

**Cleanse**

**Titanic**

Select a column to clean (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns; Alt to select all columns of the same type; Ctrl+A for all columns). Make chang... ⓘ

1 column selected

COMMIT CLEANSE CANCEL UNDO SHOW HISTORY

AUTO CLEANSING REMOVE LOW QUALITY REMOVE CORRELATED REPLACE MISSING NORMALIZATION DISCRETIZATION

Numerical missings: specific value 0.99 ✓ APPLY

Passenger Cl...	Name	Sex	Age	No of Sibling...	No of Parents...	Ticket Number	Passenger F...	Cabin	Port of
First	Fortune, Mr. Ch...	Male	19	3	2	19950	263	C23 C25 C27	Southa
First	Fortune, Mr. Mark	Male	64	1	4	19950	263	C23 C25 C27	Southa
First	Fortune, Mrs. M...	Female	60	1	4	19950	263	C23 C25 C27	Southa
First	Francatelli, Mis...	Female	30	0	0	PC 17485	56.929	E36	Cherb
First	Franklin, Mr. Th...	Male	?	0	0	113778	26.550	D34	Southa
First	Frauenthal, Dr. ...	Male	50	2	0	PC 17611	133.650	?	Southa
First	Frauenthal, Mr. ...	Male	43	1	0	17765	27.721	D40	Cherb
First	Frauenthal, Mrs...	Female	?	1	0	PC 17611	133.650	?	Southa
First	Frolicher, Miss...	Female	22	0	2	13568	49.500	B39	Cherb
First	Frolicher, Stabli...	Male	60	1	1	12567	70.200	D41	Cherb

1,309 rows - 12 columns (3 nominal, 4 numerical)

## 2. Remove Missing values from life Boat and I putt value “Nill”

**Cleanse**

**Titanic**

Select a column to clean (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns; Alt to select all columns of the same type; Ctrl+A for all columns). Make chang... ⓘ

1 column selected

COMMIT CLEANSE CANCEL UNDO SHOW HISTORY

AUTO CLEANSING REMOVE LOW QUALITY REMOVE CORRELATED REPLACE MISSING NORMALIZATION DISCRETIZATION

Nominal missings: specific value Nill ✓ APPLY

Passenger Cl...	Name	Sex	Age	No of Sibling...	No of Parents...	Ticket Number	Passenger F...	Cabin	Port of
First	Allen, Miss. Eli...	Female	29	0	0	24160	211.338	B5	Southa
First	Allison, Master...	Male	0.917	1	2	113781	151.550	C22 C26	Southa
First	Allison, Miss. H...	Female	2	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mr. Hud...	Male	30	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mrs. Hu...	Female	25	1	2	113781	151.550	C22 C26	Southa
First	Anderson, Mr. ...	Male	48	0	0	19952	26.550	E12	Southa
First	Andrews, Miss. ...	Female	63	1	0	13502	77.958	D7	Southa
First	Andrews, Mr. T...	Male	39	0	0	112050	0	A36	Southa
First	Appleton, Mrs. ...	Female	53	2	0	11769	51.479	C101	Southa

1,309 rows - 12 columns (3 nominal, 4 numerical)

### 3. Remove Missing values from Passenger Fare and I putt value “ 0.235 ”

**Cleanse**

**Titanic**

Select a column to clean (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns; Alt to select all columns of the same type; Ctrl+A for all columns). Make chang... ⓘ

1 column selected

COMMIT CLEANSE CANCEL UNDO SHOW HISTORY

Passenger Cl...	Name	Sex	Age	No of Sibling...	No of Parents...	Ticket Number	Passenger F...	Life Boat	Survive
Category	Category	Category	Number	Number	Number	Category	Number	Category	Category
First	Allen, Miss. Eli...	Female	29	0	0	24160	211.338	2	Yes
First	Allison, Master...	Male	0.917	1	2	113781	151.550	11	Yes
First	Allison, Miss. H...	Female	2	1	2	113781	151.550	Nill	No
First	Allison, Mr. Hud...	Male	30	1	2	113781	151.550	Nill	No
First	Allison, Mrs. Hu...	Female	25	1	2	113781	151.550	Nill	No
First	Anderson, Mr. ...	Male	48	0	0	19952	26.550	3	Yes
First	Andrews, Miss....	Female	63	1	0	13502	77.958	10	Yes
First	Andrews, Mr. T...	Male	39	0	0	112050	0	Nill	No
First	Appleton, Mrs. ...	Female	53	2	0	11769	51.479	D	Yes

Makes sure that all selected numerical columns are roughly on the same scale.

DISCRETIZATION

AUTO CLEANSING REMOVE LOW QUALITY REMOVE CORRELATED REPLACE MISSING NORMALIZATION DISCRETIZATION

APPLY

1,309 rows - 10 columns (6 nominal, 4 numerical)

### 4. Remove (Delete) Column “ Cabin ”.

**Transform**

**Titanic**

Select columns to transform (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns; Alt to select all columns of the same type; Ctrl+A for all columns). Make ch... ⓘ

1 column selected

COMMIT TRANSFORMATION CANCEL UNDO SHOW HISTORY

Passenger Cl...	Name	Sex	Age	No of Sibling...	No of Parents...	Ticket Number	Passenger F...	Cabin	Port of
Category	Category	Category	Number	Number	Number	Category	Number	Category	Category
First	Allen, Miss. Eli...	Female	29	0	0	24160	211.338	B5	Southa
First	Allison, Master...	Male	0.917	1	2	113781	151.550	C22 C26	Southa
First	Allison, Miss. H...	Female	2	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mr. Hud...	Male	30	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mrs. Hu...	Female	25	1	2	113781	151.550	C22 C26	Southa
First	Anderson, Mr. ...	Male	48	0	0	19952	26.550	E12	Southa
First	Andrews, Miss....	Female	63	1	0	13502	77.958	D7	Southa
First	Andrews, Mr. T...	Male	39	0	0	112050	0	A36	Southa
First	Appleton, Mrs. ...	Female	53	2	0	11769	51.479	C101	Southa

RENAMERENAME CHANGE TYPE REMOVE COPY FILTER RANGE SAMPLE SORT

APPLY Deletes all selected columns from the data set.

1,309 rows - 12 columns (8 nominal, 4 numerical)

## 5. Remove (Delete) Column “ Port of Embarkment ”.

The screenshot shows the RapidMiner Studio interface in the 'Turbo Prep' view. On the left, a sidebar lists various transformation operations: RENAME, CHANGE TYPE, REMOVE (which is currently selected), COPY, FILTER, RANGE, SAMPLE, and SORT. The main area displays the 'Titanic' dataset with 1,309 rows and 11 columns. The 'Port of Emb...' column is highlighted in orange. At the bottom of the transformation interface, there are buttons for 'COMMIT TRANSFORMATION' and 'CANCEL', along with 'UNDO' and 'SHOW HISTORY' buttons. The status bar at the bottom right indicates the time as 11:32 AM and the date as 11/16/2020.

## 6. Final Output Result Window.

The screenshot shows the RapidMiner Studio interface in the 'Process' tab. The central workspace displays a data pipeline. It starts with a 'Retrieve' operator ('Loading Titanic') connected to a 'Subprocess' operator. This is followed by four more 'Subprocess' operators labeled '(2)', '(3)', '(4)', and '(5)'. The final operator is 'Select Attributes'. To the left of the process, the 'Repository' pane shows the local repository with a file named 'Answer\_01.rmp'. Below it, the 'Operators' pane lists categories like Data Access, Blending, Cleansing, Modeling, Scoring, Validation, and Utility. To the right, the 'Parameters' pane shows parameters for the process, including 'verbosity: init' and 'logfile'. The 'Help' pane provides information about the 'Process' operator. The status bar at the bottom right shows the time as 11:40 AM and the date as 11/16/2020.

Question no:- 1.2 :- Create a process to find out the total Passenger fare paid by females belonging to age more than 30 for the first passenger class?

Ans:- Procedure & Solution:-

- \* first is opened T-SQL prep and transform the Age column
- filter the column according condition then I filtered the column sex according condition
- \* Select the female.
- \* Transform the column Passenger class and filtered the second class. and last step create process and export the process.

End the Solution  
1.2 Question  
Total sum

## 1. Filter the age Column According to Condition age > 30.

Views: Design Results Turbo Prep Auto Model Deployments Find data, operators... etc All Studio

Transform

Titanic (2)

1 column selected

RENAME

CHANGE TYPE

REMOVE

COPY

FILTER

>

30

✓ APPLY

RANGE

COMMIT TRANSFORMATION CANCEL UNDO SHOW HISTORY

Passenger Cl... Category	Name Category	Sex Category	Age Number	No of Sibling... Number	No of Parents... Number	Ticket Number Category	Passenger F... Number	Cabin Category	Port of Category
First	Allen, Miss. Eli...	Female	29	0	0	24160	211.338	B5	Southa
First	Allison, Master....	Male	0.917	1	2	113781	151.550	C22 C26	Southa
First	Allison, Miss. H...	Female	2	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mr. Hud...	Male	30	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mrs. Hu...	Female	25	1	2	113781	151.550	C22 C26	Southa
First	Anderson, Mr. ...	Male	48	0	0	19952	26.550	E12	Southa
First	Andrews, Miss....	Female	63	1	0	13502	77.958	D7	Southa
First	Andrews, Mr. T...	Male	39	0	0	112050	0	A36	Southa
First	Appleton, Mrs. ...	Female	53	2	0	11769	51.479	C101	Southa

1,309 rows - 12 columns (8 nominal, 4 numerical)

## 2. Filter Column Sex According to Condition Sex == Female.

Views: Design Results Turbo Prep Auto Model Deployments Find data, operators... etc All Studio

Transform

Titanic (2)

1 column selected

RENAME

CHANGE TYPE

REMOVE

COPY

FILTER

equals

Female

✓ APPLY

RANGE

COMMIT TRANSFORMATION CANCEL UNDO SHOW HISTORY

Passenger Cl... Category	Name Category	Sex Category	Age Number	No of Sibling... Number	No of Parents... Number	Ticket Number Category	Passenger F... Number	Cabin Category	Port of Category
First	Anderson, Mr. ...	Male	48	0	0	19952	26.550	E12	Southa
First	Andrews, Miss....	Female	63	1	0	13502	77.958	D7	Southa
First	Andrews, Mr. T...	Male	39	0	0	112050	0	A36	Southa
First	Appleton, Mrs. ...	Female	53	2	0	11769	51.479	C101	Southa
First	Arfagaveftia, Mr...	Male	71	0	0	PC 17609	49.504	?	Cherb
First	Astor, Col. Joh...	Male	47	1	0	PC 17757	227.525	C62 C64	Cherb
First	Barkworth, Mr. ...	Male	80	0	0	27042	30	A23	Southa
First	Baxter, Mrs. Ja...	Female	50	0	1	PC 17558	247.521	B58 B60	Cherb
First	Bazzani, Miss. ...	Female	32	0	0	11813	76.292	D15	Cherb

437 rows - 12 columns (8 nominal, 4 numerical)

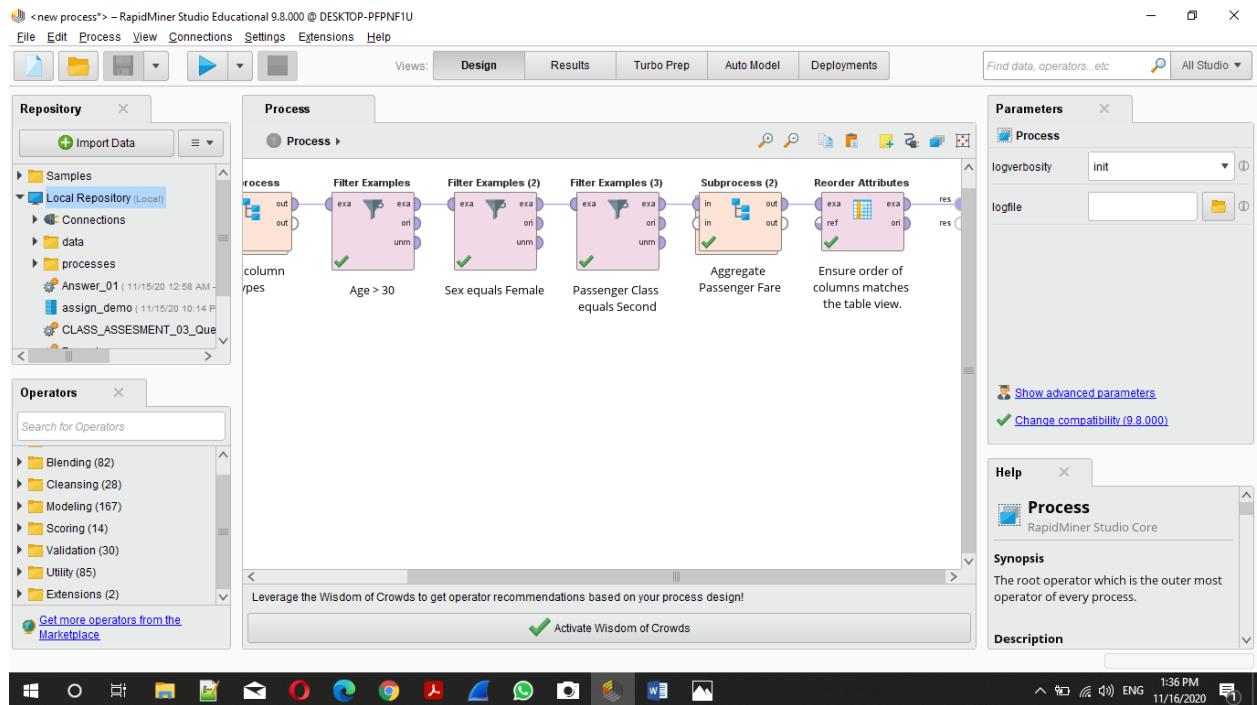
### 3. Filter Column Sex According to Condition Passenger Class == “Second”.

The screenshot shows the RapidMiner Studio interface in the Turbo Prep view. On the left, there's a sidebar with options like RENAME, CHANGE TYPE, REMOVE, COPY, FILTER, and RANGE. The FILTER section has a dropdown set to 'equals' and a text input field containing 'Second', with a '✓ APPLY' button. The main area displays a table titled 'Titanic (2)' with 153 rows and 12 columns. The first column is 'Passenger Class' with categories like 'Second', 'Cabin', and 'Port'. The second column is 'Name', third is 'Sex', fourth is 'Age', fifth is 'No of Siblins...', sixth is 'No of Parents...', seventh is 'Ticket Num...', eighth is 'Passenger ...', ninth is 'Cabin', tenth is 'Port', and eleventh is 'Survived'. The last column is 'Pclass'. The table header includes 'Category' under 'Passenger Class' and 'Number' under 'Age', 'No of Siblins...', 'No of Parents...', 'Ticket Num...', 'Passenger ...', 'Cabin', and 'Port'. The bottom right of the table shows '153 rows - 12 columns (8 nominal, 4 numerical)'. The system status bar at the bottom right indicates '11:53 AM 11/16/2020'.

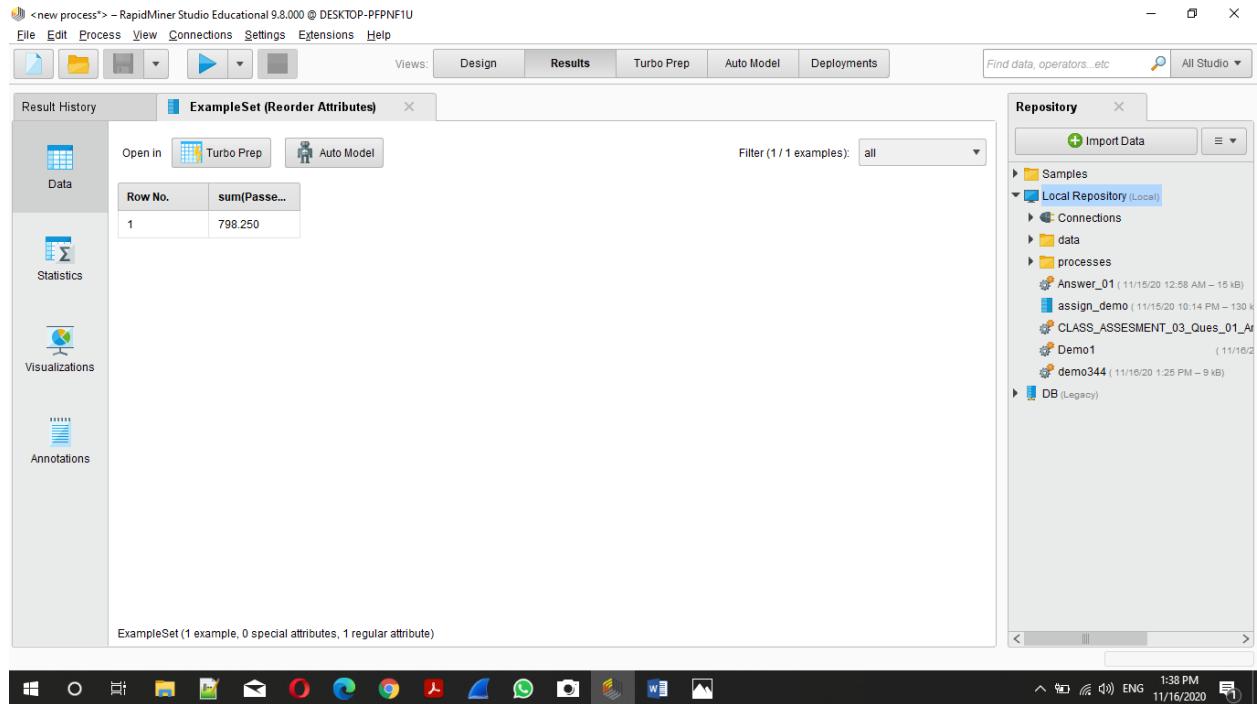
### 4. Use Aggregate Function for the Total sum of passenger Fare

The screenshot shows the RapidMiner Studio interface in the Turbo Prep view. On the left, there's a sidebar with various columns listed: 'Number', 'No of Siblings or Spouses on Board', 'No of Parents or Children on Board', 'Ticket Number', 'Passenger Fare', 'Cabin', 'Port of Embarkation', 'Life Boat', and 'Survived'. The 'Passenger Fare' column is selected for grouping. The main area displays a table titled 'Titanic (2)' with a single row showing the aggregate result. The result is a green box labeled 'Passenger Fare' with 'sum' and '798.250'. Below the table, it says '1 Aggregate 1 rows - 1 columns (1 numerical)'. The system status bar at the bottom right indicates '1:24 PM 11/16/2020'.

## 5. Final Design.



## 6. Output :-



Question No:- 1.3:- Create a process to find the maximum age of male and female having 1 sibling belonging to second passenger class.

Solution:-

first I load the Titantic data set in Turbo prep. first is filtered the data of passenger class according to condition.

- \* Second filtered the data of No of siblings
- \* Use aggregate function and put the column of age and I find the maximum value after that I create the process and

Export the Process .

Thank you.  
End of Solution of question 3.  
Talbut Shar

## 1. Filter Column No of Siblings According to Condition, No of Siblings == 1.

**Transform**

**Titanic (3)**

1 column selected

RENAME  
CHANGE TYPE  
REMOVE  
COPY  
**FILTER**

=  
1  
✓ APPLY  
RANGE

Passenger Cl... Category	Name Category	Sex Category	Age Number	No of Sibling... Number	No of Parents... Number	Ticket Number Category	Passenger F... Number	Cabin Category	Port of Category
First	Allen, Miss. Eli...	Female	29	0	0	24160	211.338	B5	Southa
First	Allison, Master...	Male	0.917	1	2	113781	151.550	C22 C26	Southa
First	Allison, Miss. H...	Female	2	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mr. Hud...	Male	30	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mrs. Hu...	Female	25	1	2	113781	151.550	C22 C26	Southa
First	Anderson, Mr. ...	Male	48	0	0	19952	26.550	E12	Southa
First	Andrews, Miss....	Female	63	1	0	13502	77.958	D7	Southa
First	Andrews, Mr. T...	Male	39	0	0	112050	0	A36	Southa
First	Appleton, Mrs. ...	Female	53	2	0	11769	51.479	C101	Southa

1,309 rows - 12 columns (8 nominal, 4 numerical)

## 2. Filter Column Passenger Class == “second” According to Condition.

**Transform**

**Titanic (3)**

1 column selected

RENAME  
CHANGE TYPE  
REMOVE  
COPY  
**FILTER**

equals  
Second  
✓ APPLY  
RANGE

Passenger Cl... Category	Name Category	Sex Category	Age Number	No of Sibling... Number	No of Parents... Number	Ticket Number Category	Passenger F... Number	Cabin Category	Port of Category
First	Allison, Master...	Male	0.917	1	2	113781	151.550	C22 C26	Southa
First	Allison, Miss. H...	Female	2	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mr. Hud...	Male	30	1	2	113781	151.550	C22 C26	Southa
First	Allison, Mrs. Hu...	Female	25	1	2	113781	151.550	C22 C26	Southa
First	Andrews, Miss....	Female	63	1	0	13502	77.958	D7	Southa
First	Astor, Col. Joh...	Male	47	1	0	PC 17757	227.525	C62 C64	Cherbo
First	Astor, Mrs. Joh...	Female	18	1	0	PC 17757	227.525	C62 C64	Cherbo
First	Beckwith, Mr. R...	Male	37	1	1	11751	52.554	D35	Southa
First	Beckwith, Mrs. ...	Female	47	1	1	11751	52.554	D35	Southa

319 rows - 12 columns (8 nominal, 4 numerical)

### 3. Maximum age of Male and Female .

The screenshot shows the RapidMiner Studio interface with the "Turbo Prep" tab selected. In the center, there is a "Pivot" tool titled "Titanic (4)". A message says "Drag columns from the left to build a pivot table." Below this are buttons for "COMMIT PIVOT", "CLEAR TABLE", and "CANCEL". On the left, a list of columns is shown under "Group-By": Passenger Class, Name, Sex, Age, No of Siblings or Spouses on Board, No of Parents or Children on Board, Ticket Number, Passenger Fare, and Cabin. On the right, a "Column Grouping" section shows an aggregate for "Age" with the value "maximum" set to "63". Below this, it says "1 Aggregate" and "1 rows - 1 columns (1 numerical)". At the bottom, there are tabs for "Aggregates" and "Age maximum". The Windows taskbar at the bottom shows various application icons.

### 4. Final Design.

The screenshot shows the RapidMiner Studio interface with the "Design" tab selected. The central workspace displays a process flow for "Loading Titanic (3)". The flow starts with a "Retrieve" operator (purple) connected to a "Subprocess" operator (orange). The "Subprocess" operator has two parallel paths: one for "Unify column types" and another for "Filter Examples (1)" with the condition "No of Siblings or Spouses on Board = 1". These are followed by another "Subprocess" operator (orange) with two parallel paths: one for "Passenger Class equals Second" and another for "Aggregate Age". Finally, there is a "Reorder Attributes" operator (pink) with the condition "Ensure order of columns matches the table view". To the left, the "Repository" panel shows the "Local Repository (Local)" with "Samples" and "processes" listed. The "Operators" panel lists categories like Blending, Cleansing, Modeling, Scoring, Validation, Utility, and Extensions. To the right, the "Parameters" panel shows "Process" parameters: "verbosity" set to "init" and "logfile" set to an empty field. The "Help" panel provides information about the "Process" operator. The Windows taskbar at the bottom shows various application icons.

## 5. Output :-

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. The main area displays an 'ExampleSet (Reorder Attributes)' with the following data:

Row No.	maximum(A...)
1	63

The left sidebar has tabs for Data, Statistics, Visualizations, and Annotations. The right sidebar shows the 'Repository' with a tree view of local files and processes, including 'Answer\_01', 'assign\_demo', 'CLASS\_ASSESSMENT\_03\_Ques\_01\_A...', 'Demo1', 'demo344', and 'max'. The system tray at the bottom shows the date and time as 11/16/2020 1:44 PM.

End Solution Of Question No 01

Q2) You are given a csv file named "Whales with Porpoises.csv". You have to split the data based on the given condition and later on count the number of rows in each category using ssis

Output name	Condition
Harbour Species	Species of Harbour type
Other Species	All other species

Question No:- 02:- You are given a csv file named "Whales with Porpoise.csv". You have to split the data based on the given condition and later on count the number of rows in each category using ssis.

→ Output Name	Condition
1. Harbour Species	Species of Harbour type
2. Other Species	All other Type .

Solution:- first I opened visual studio create new project give the name of project. go to ssis Tool box select Data flow Task, click on data flow & then It will open Data flow window select Flat file source from source and configure it and load the CSV file after that select the column according to condition.

→ take the conditional split and connect & configure it the column specie  
1. Harbour Species == "Harbour ~~specie~~ Porpoise".  
2. Other Species != "Harbour ~~Porpoise~~".

→ ~~then~~ make the connection between Aggregate function. take two aggregate function .

(2)

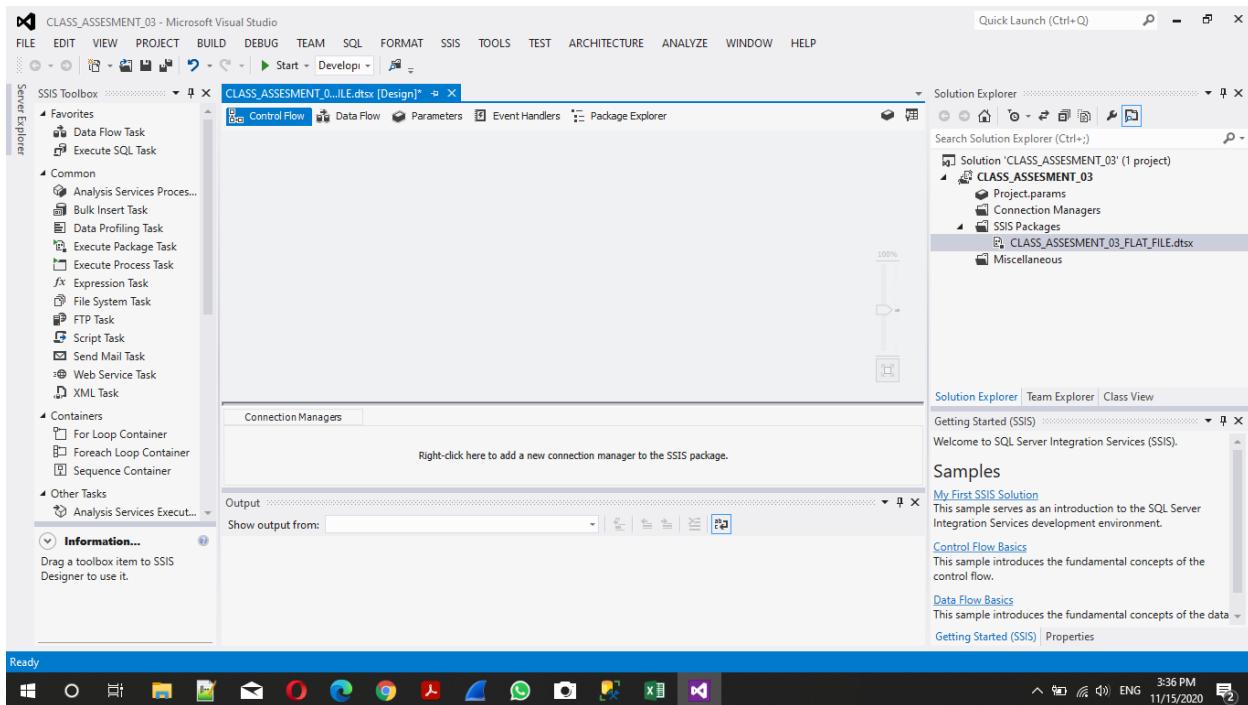
16/11/2020

first for Harbour Species and second per other species, for the count  
the same and make the connection from conditional split  
Select the column name which you want to count. it will generate ~~the~~  
~~same~~ count of rows.

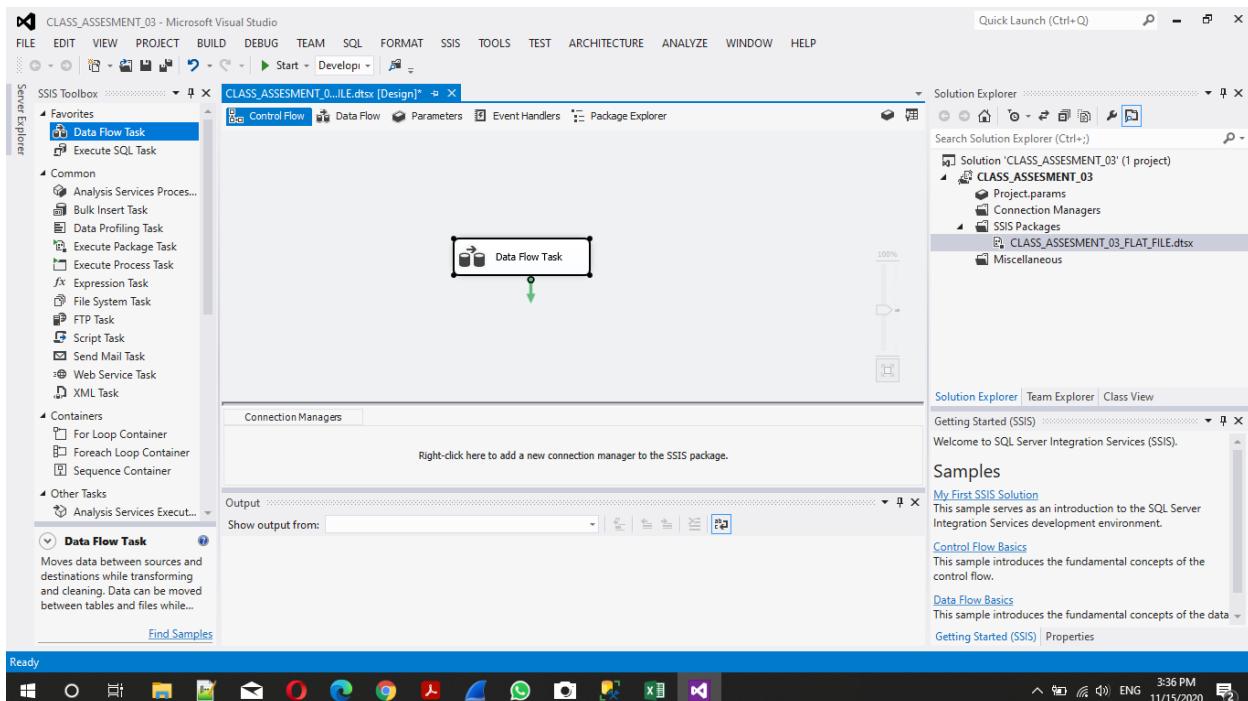
- take two Flat file Destination and make the connection from Aggregate and configure the destination file. make the connection where you store the data. create a new file for the store the data like Harbour species for One file and other species for another file then use delimiter accordingly if you like:-, Tab etc. and most important Map it Properly.
- click the start button for the process/result/output.

Talebul Islam.  
16/11/2020.  
DOMS

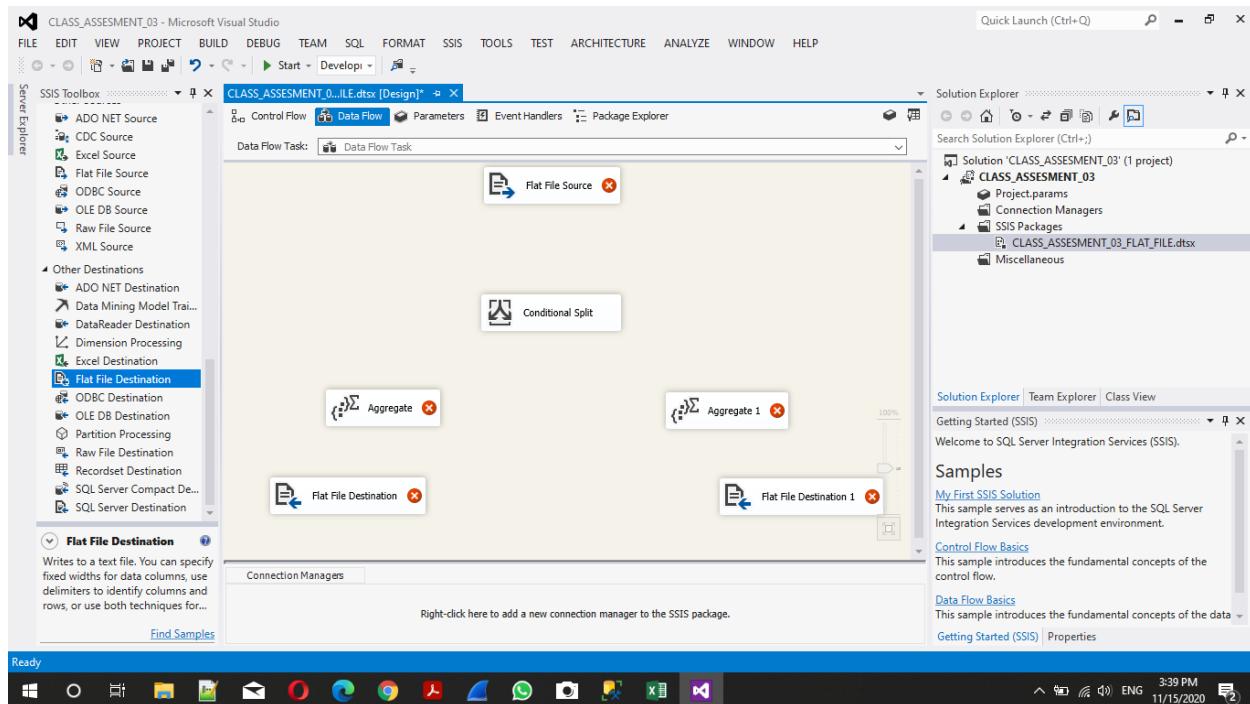
## 1. Opened visual Studio



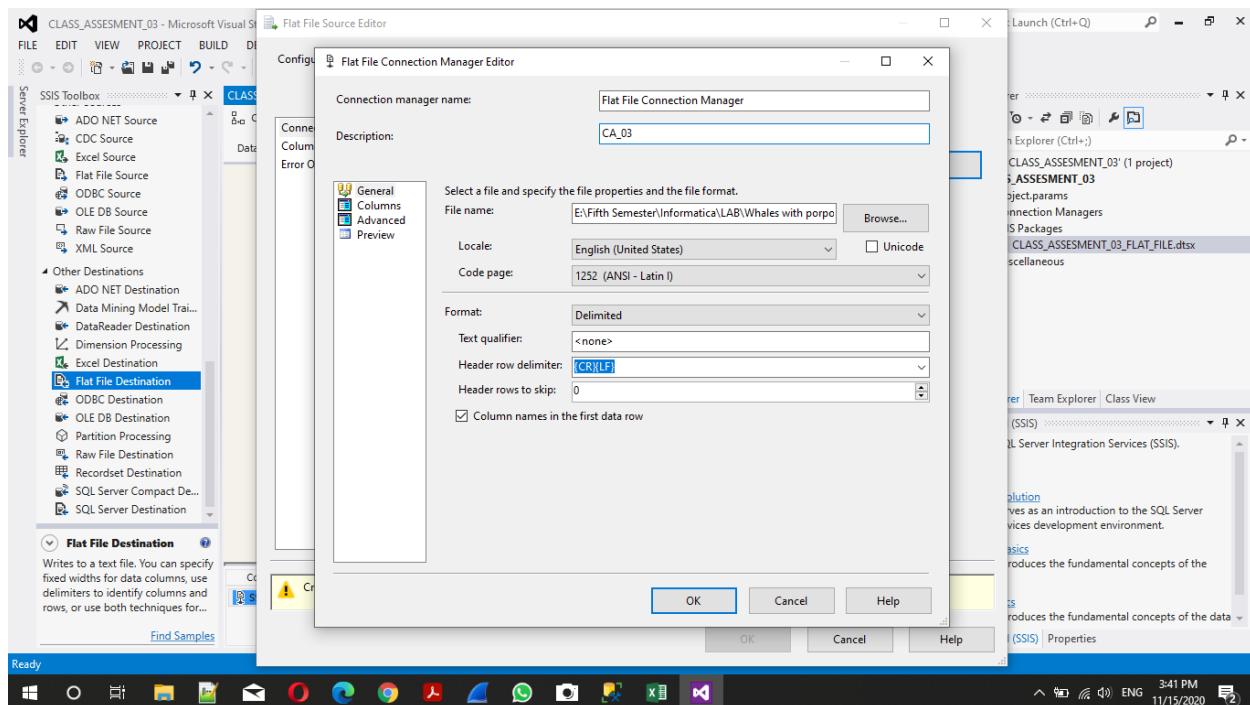
## 2. Select Data Flow Task From SSIS Toolbox



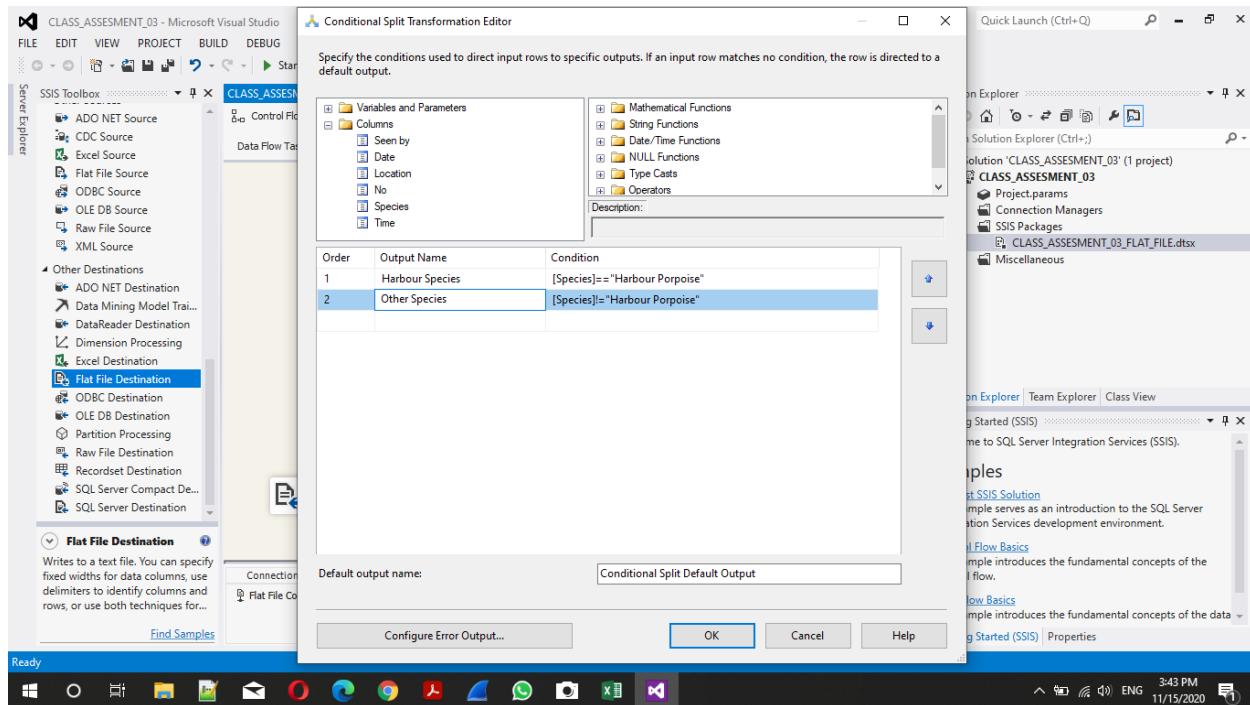
### 3. Select All The Required Tool From SSIS Toolbox. Like: - Source file, Conditional Split, Aggregate Tool, Destination File.



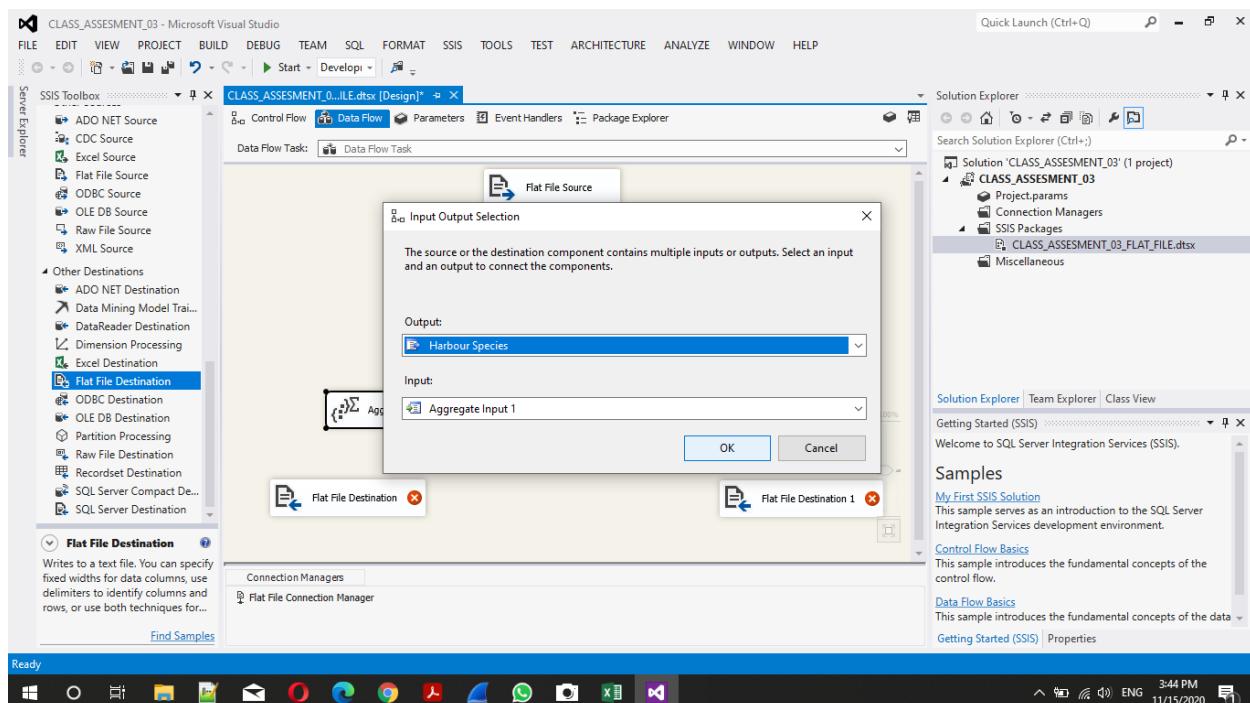
### 4. Configure Source Flat File.



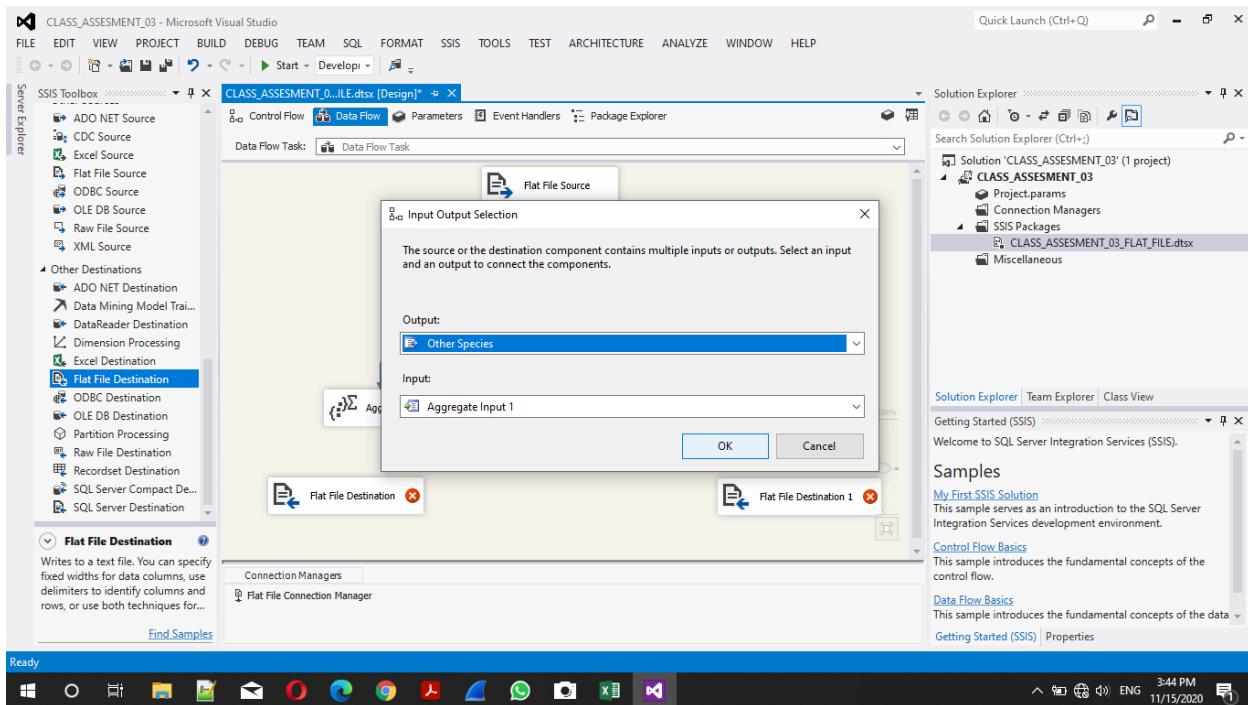
## 5. Configure Conditional Split According to Two Condition.



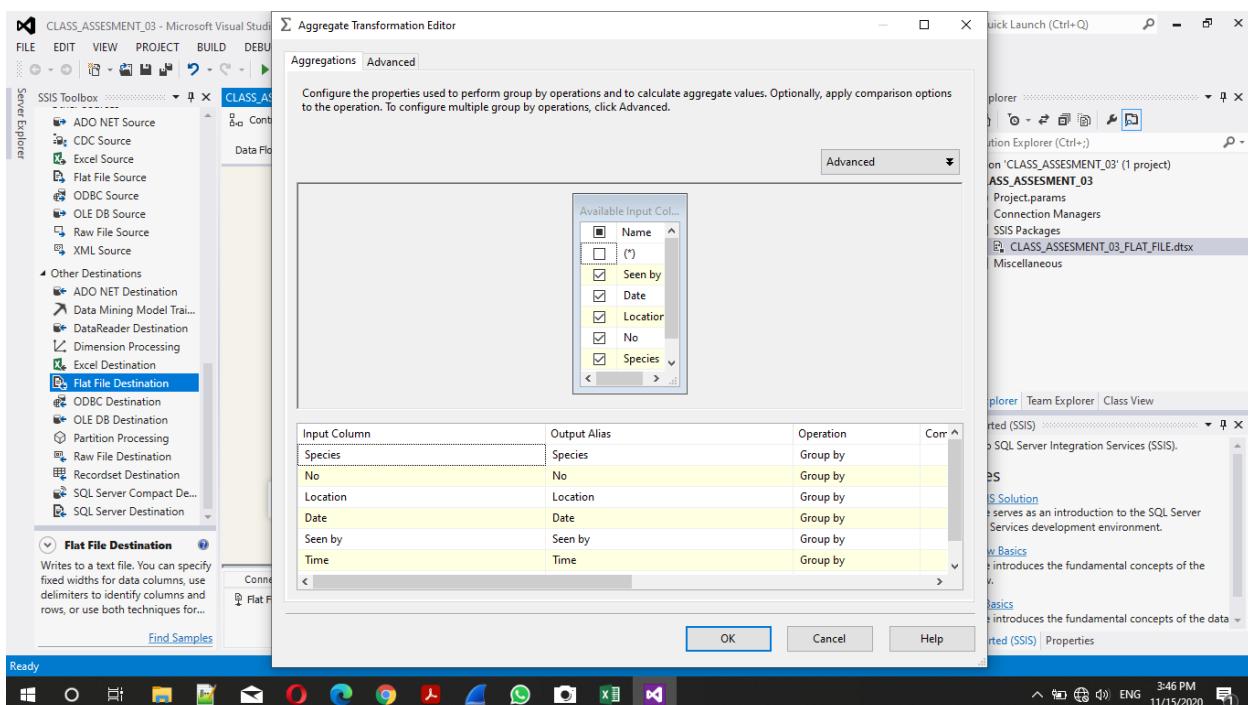
## 6. Connection of Harbour Species Aggregate Function Tool For rows count.



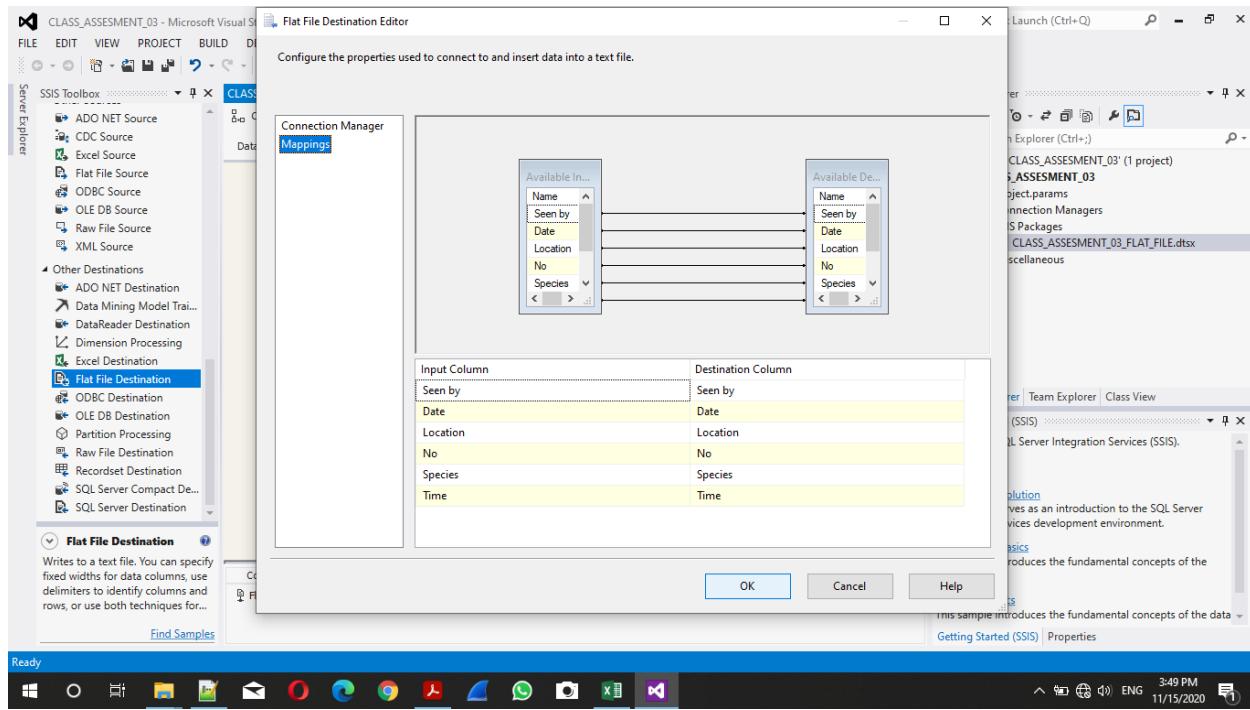
## 7. Connection of Other Species Aggregate Function Tool for Row Count.



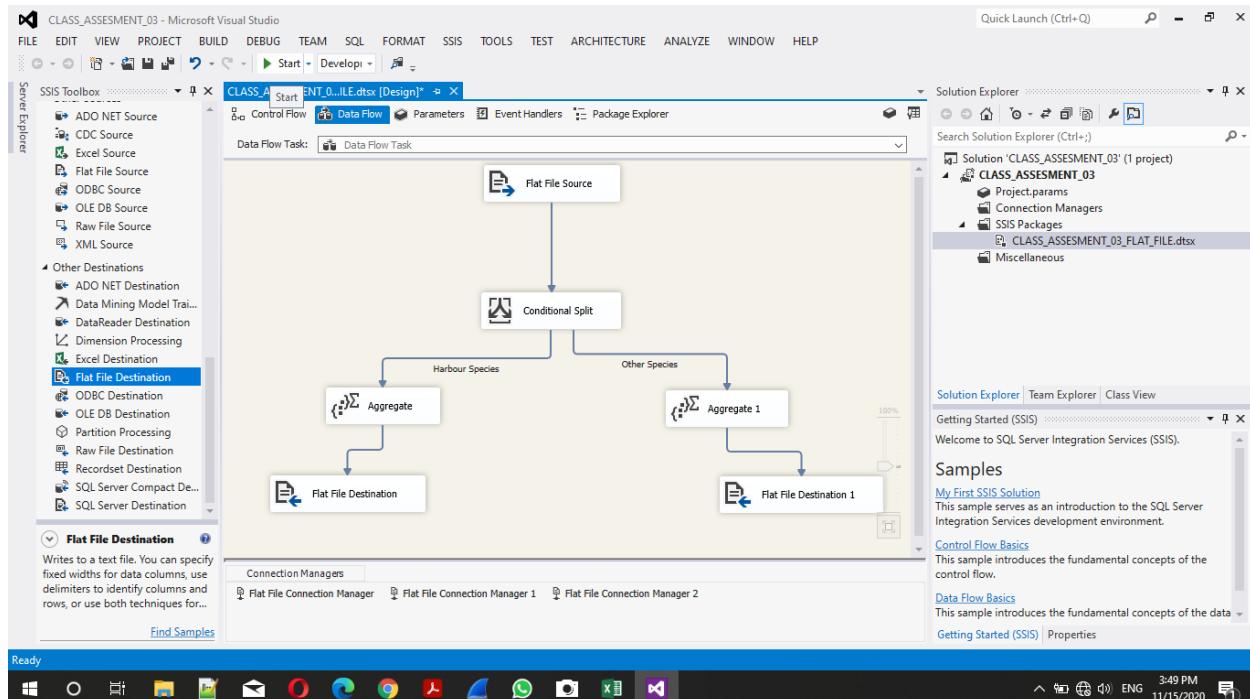
## 8. Configure Aggregate Function.



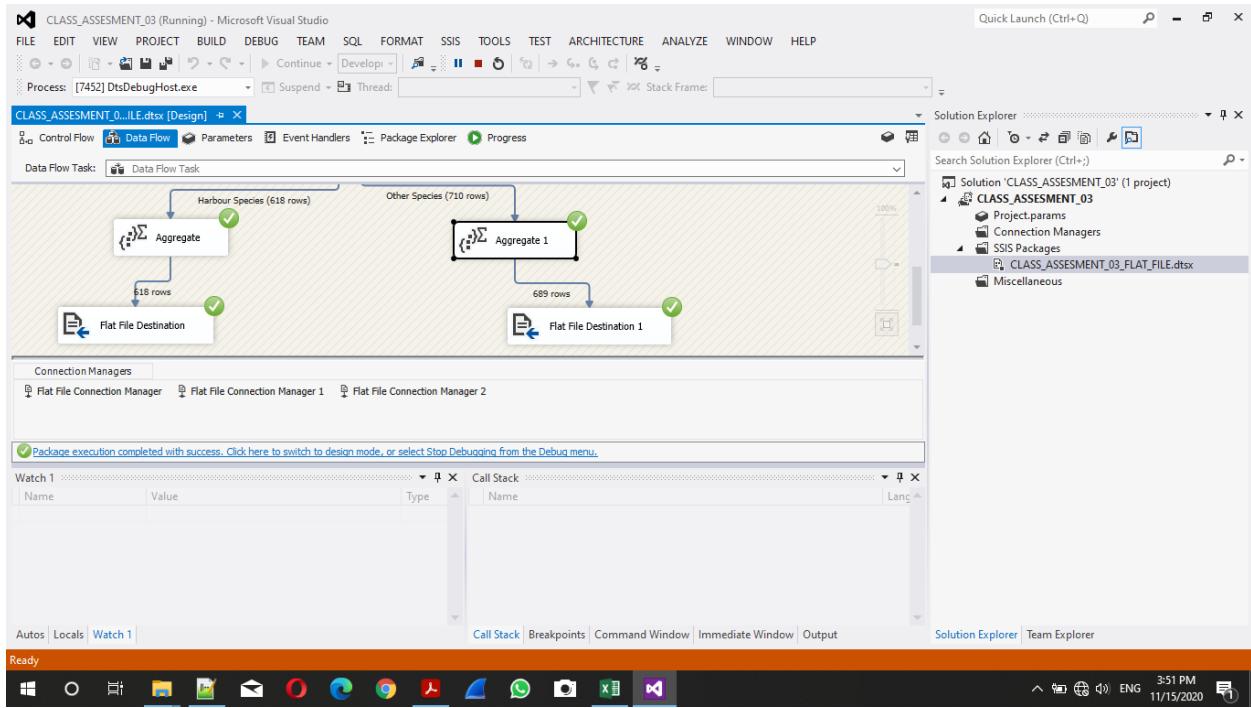
## 9. Mapping of flat File to destination Flat file.



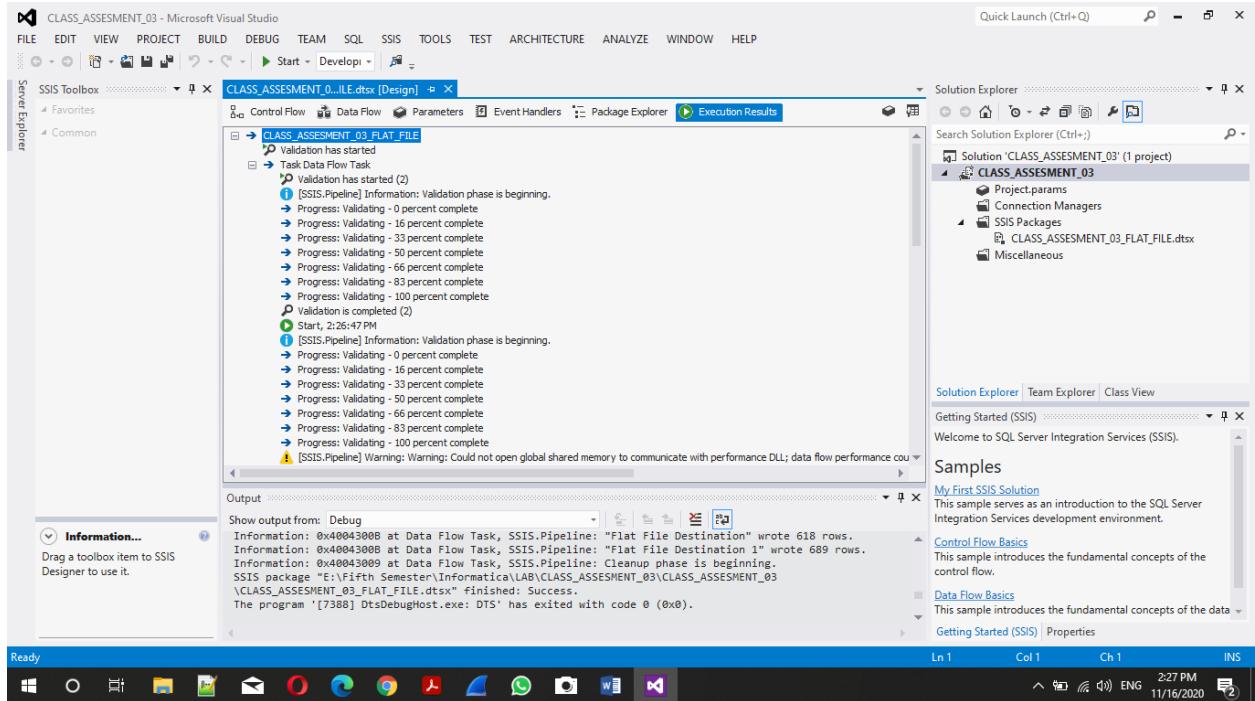
## 10.Final Data flow Design



## 11. Output Window.



## 12. Execution Result.



-----+THANK YOU MA'AM++-