MODUL 10
PROJECT PLAN ON THE LINES OF DCOVA&I

THEME DESCRIPTION

The student is able to perform sanity checks to be performed on data, typically, it may involve any or all of the following:

- Checking whether the appropriate numbers of rows and columns/observations and variables have been imported into the analytics system. You can check this against the source system.
- Checking the formats of the various variables in the data set.
- Checking for missing data.

WEEKLY LEARNING OUTCOMES (SUB-LESSONS)

CLO-1-Sub-CLO-10:
Able to implement the early stages of data science implementation including business opportunities as well as analysis and selection of effective data variables– C4.

You are asked to create a project plan along the lines of DCOVA and I (**D**efine, **C**ollect, **O**rganize, **V**isualize, **A**nalyze, and **I**nsights process).

1. First you **define** the problem. Clean the data and create a project datamart to do analytics for Resolution Time, in other words, the time, in a number of days, taken to resolve the service request (to be calculated as the difference between SR Close Date and SR Open Date).
   - ➲ Create the y variable.
2. **Collect** the relevant data. The data for the project comes from one file, CaseStudy1.csv.
3. **Organize** the data. Manipulate data, create derived variables through calculation, and understand the missing values.
4. **Visualize** the data.
   - ➲ Univariate analysis of y
   - ➲ Multivariate analysis: correlation
5. Get **Insights**.
   - ➲ Drop the variables used to create y since these variables have been used to create the business problem.

PRACTICUM SUPPORT

a. Windows Operating System
b. .NET Framework (*installed*)
c. SAS® Enterprise Guide version 8.3 (8.3.0.103) (32-bit) ODA (*installed*)

CASE STUDY

An UMN-enabled services (UMNES) company wants to understand its data related to service requests for customers. These requests are related to particular products that the company manufactures. The requests can be divided into three priorities: Low, Medium, and High. Each of these priority levels has a different service level agreement (SLA) for resolution.
The data consists of the following fields:

- ServiceRequestNo: This is the unique ID for each service request logged on the system.
- ServiceRequestStatus: As of a certain date, this is the status of the service request.
- TypeOfEnagagement: This is the type of work the company is doing for the client.
- Incident/Problem: This specifies whether the service request is a problem.
- SR Priority: This is the priority level of the service request.
- SR Open Date: This is the date on which the service request was logged in the system.
- SR Close Date: This is the date on which the service request was closed in the system.
- Product : This is the name of the product for which the service request was raised.
- Geography: This is the continent in which the client belongs.
- Country: This is the country to which the client belongs.

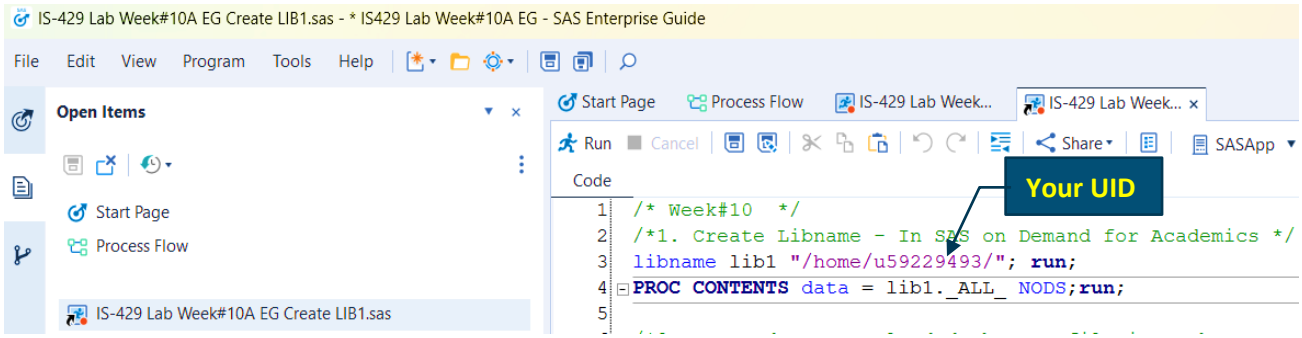
PRACTICUM STEPS

1. **Introduction to SAS® Enterprise Guide**

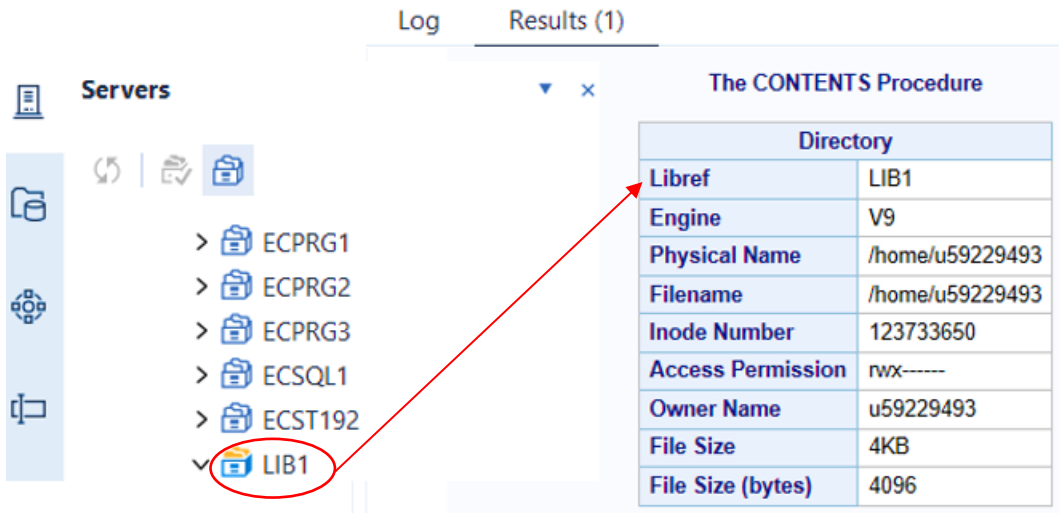
   Read and learn the Introduction to **SAS® Enterprise Guide.pdf** document to understand the SAS® Enterprise Guide (**EG**) software along with the installation and usage steps

2. **Create a Libname**

   - In SAS OnDemand for Academics, the library is preset for the course.
   - You can view the name via the SAS Studio link. (Hint: Google is your best friend when you want to find relevant documents and do-it-yourself processes. SAS has extensive online help).

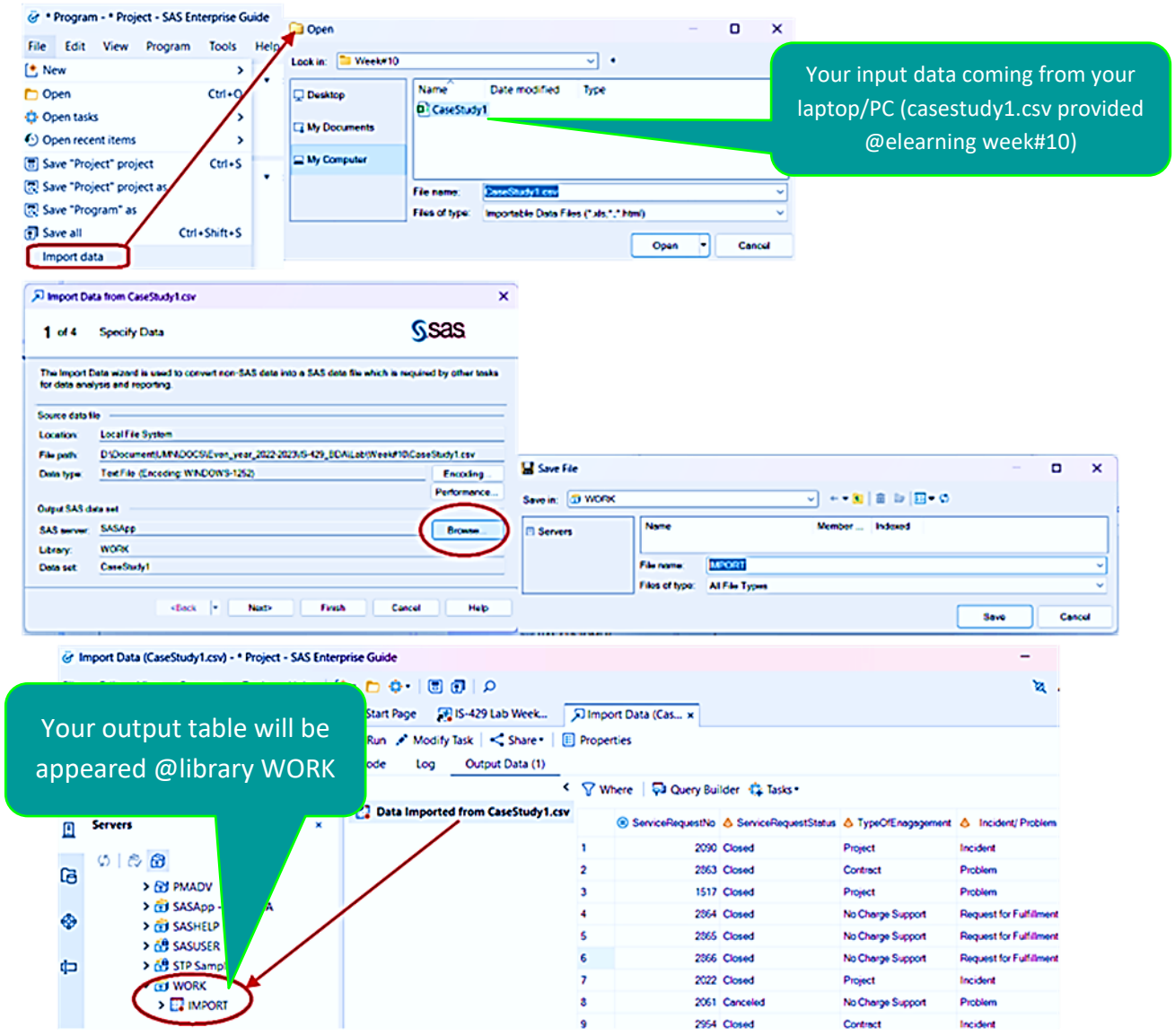


   a. Start, SAS® Enterprise Guide and type it on the above code (line 3 and 4)
   b. You will get the result of your library creation as follows:



3. Import the data into SAS cloud directory

- Loaded the CSV file into the SAS on-demand cloud directory.
- Colect the relevant data: - The data for the project comes from one file "CaseStudy1.csv".
- An easier way to import without touching the cloud-based server is to use the button-driven option
- File ➤ Import Data. This will open an Import Wizard through which you can pull the data into the SAS system from your desktop.
- Please do follow these instructions below:



- Did the file get loaded? Is all the data present?
- PROC CONTENTS will help you determine this.

a. To help you determine the result, please type it in the code below.

b. Still using the program tray in the same EG, add the following statement below it and run it.

```
6  /*2.Import data : Lloaded the csv file into the SAS
7  PROC CONTENTS DATA=WORK.IMPORT; RUN;
```

c. Screenshot of the results of the code above matches the output **figure A** into the word file "**IS-429 Lab W#10 yourname (doc/docx)**".

4. Store the SAS data set back in the SAS cloud library

- In this case, you are asked to save CaseStudy1 data from library.WORK to library.LIB1 with table name CS1.

a. Still using the program tray in the same EG, add the following statement below it and run it.

```
 9  /* 3. Store the SAS dataset back in SAS on Demand library
10 DATA LIB1.CS1;
11  SET WORK.IMPORT; RUN;
```

b. Screenshot of the results of the code above matches the output **figure B** into the previous word file "**IS-429 Lab W#10 yourname (doc/docx)**".

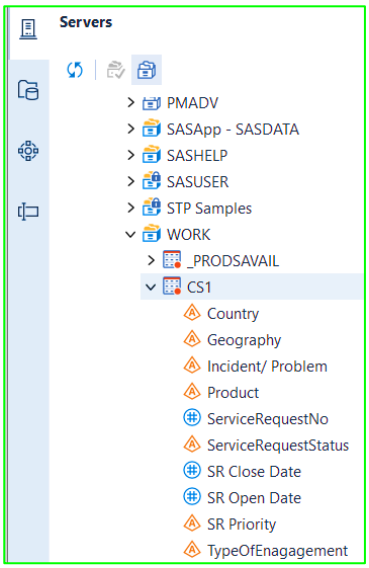
5. Create a copy of the file WORK.IMPORT as WORK.CS1

- Work on the file WORK.IMPORT in the WORK directory.
- Create a copy of the file WORK.IMPORT as WORK.CS1.

a. Still using the program tray in the same EG, add the following statement below it and run it.

```
18  Create copy of file WORK.IMPORT as WORK.CS1. */
19 DATA WORK.CS1;
20  SET WORK.IMPORT; RUN;
```

b. The output of the code above produces the following result:



6. Define the problem: Clean the data to resolve the SR

- Clean the data and create a Project Datamart to do Analytics for Resolution Time i.e., Time, in number of days, taken to **resolve the Service Request/SR** (to be calculated as difference between SR Close Date and SR Open Date.)
- Create the **y variable of Resolution Time**, which is the difference between SR Close Date and SR Open Date.
-

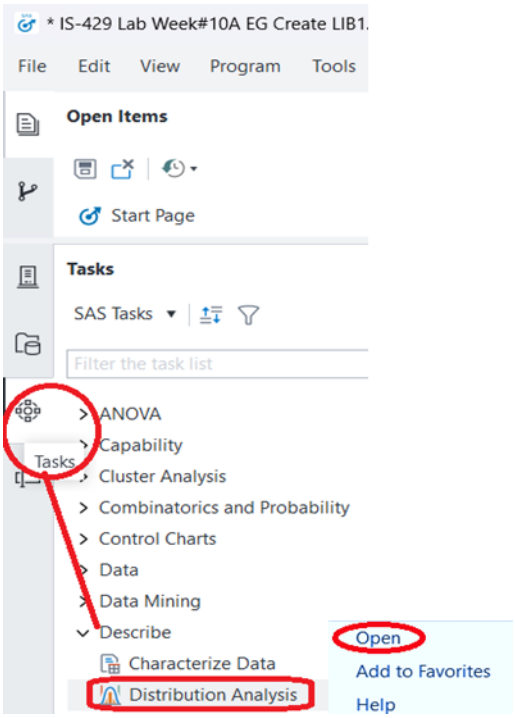
a. Still using the program tray in the same EG, add the following statement below it and run it.

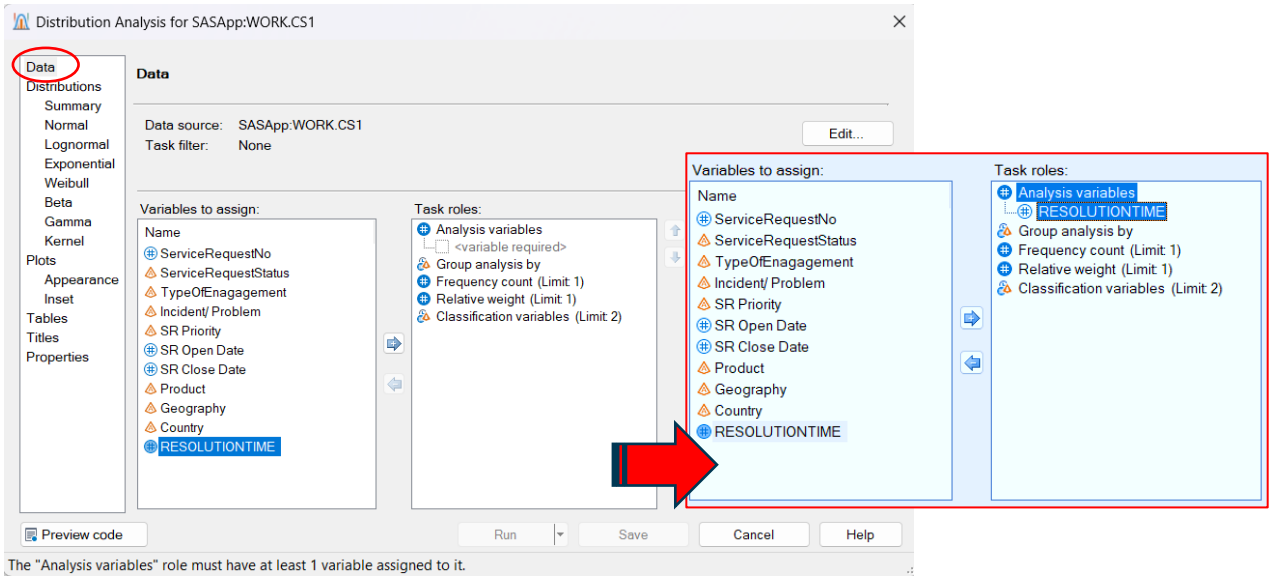
b. Screenshot of the results of the code above matches the output **figure C** into the previous word file "**IS-429 Lab W#10 yourname (doc/docx)**".

7. Analyze, Visualize, and get Insights: Understand the distribution of the y variable

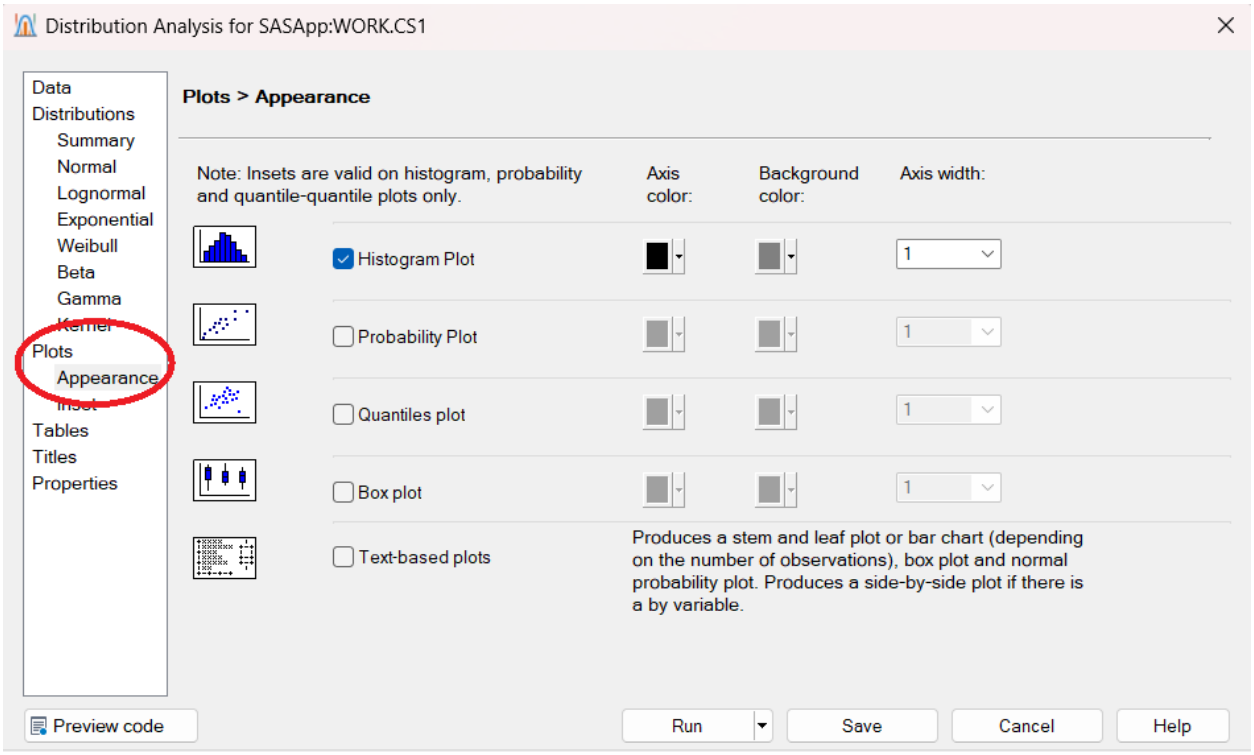
- Organize the data. Manipulate data, create derived variables through calculation, understand the missing values and Visualize the data with Univariate analysis of y.
- Let's understand the distribution of the y variable.
- Choose the Tasks > tab and then the Describe data tab and then choose Distribution Analysis.
- Choose Normal Distribution and Histogram Plot.

a. Please do follow these instructions below to create the Distribution Analysis:

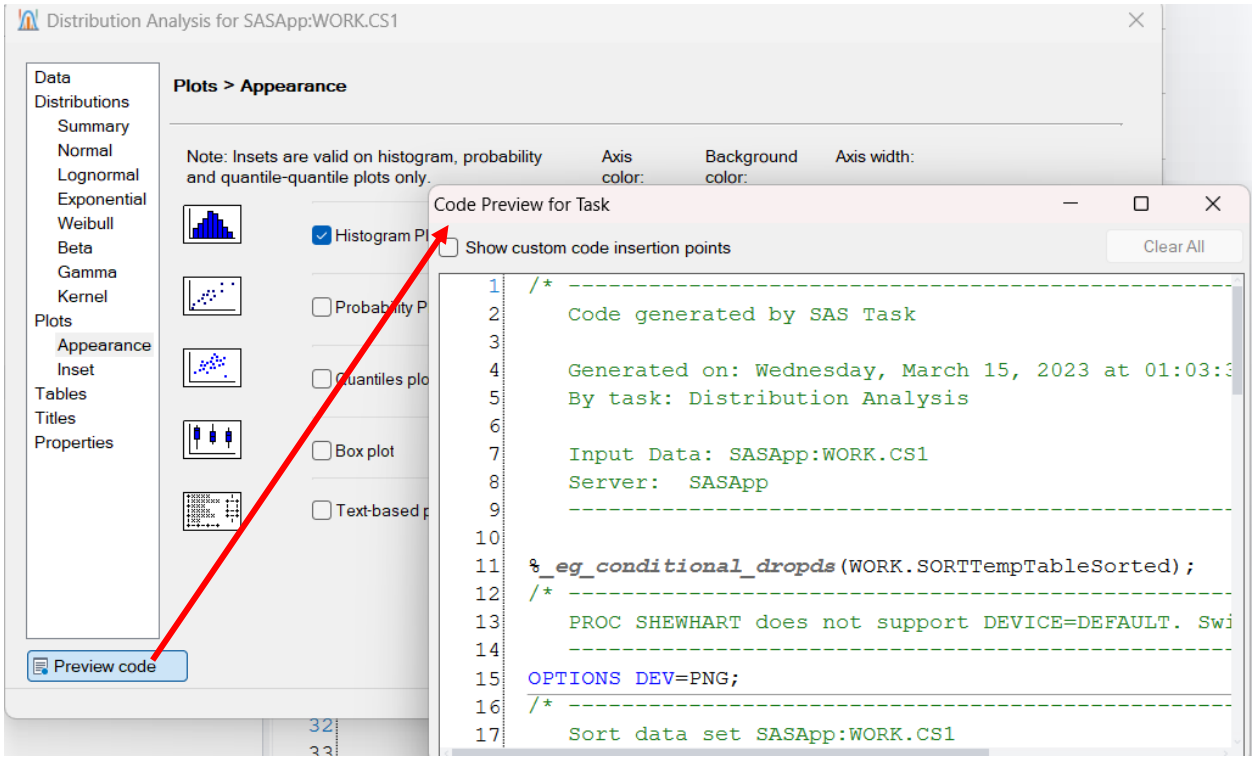




b. **Visualize**: Do follow these instructions below to create the Normal Distribution and Histogram Plot:



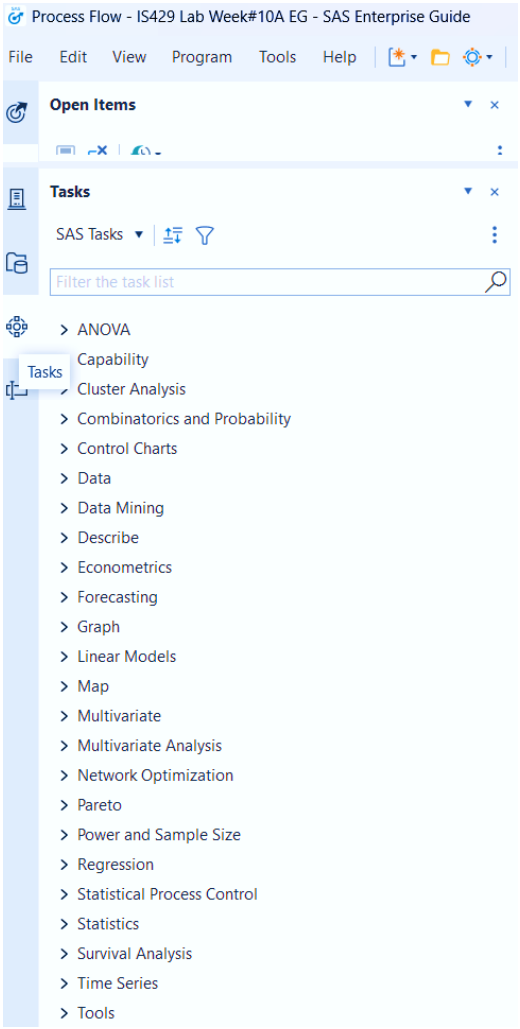
c. The EG system will share the program code for you for free. Click on Preview Code to see the SAS code for the options you have choosed thru the button menu !!



d. Screenshot of the results of the code above matches the output **figure D** into the previous word file "**IS-429 Lab W#10 yourname (doc/docx)**".

8. Process Flow

- Then you can do analysis, visualization and get other new insights through many features of the tasks provided at EG below.



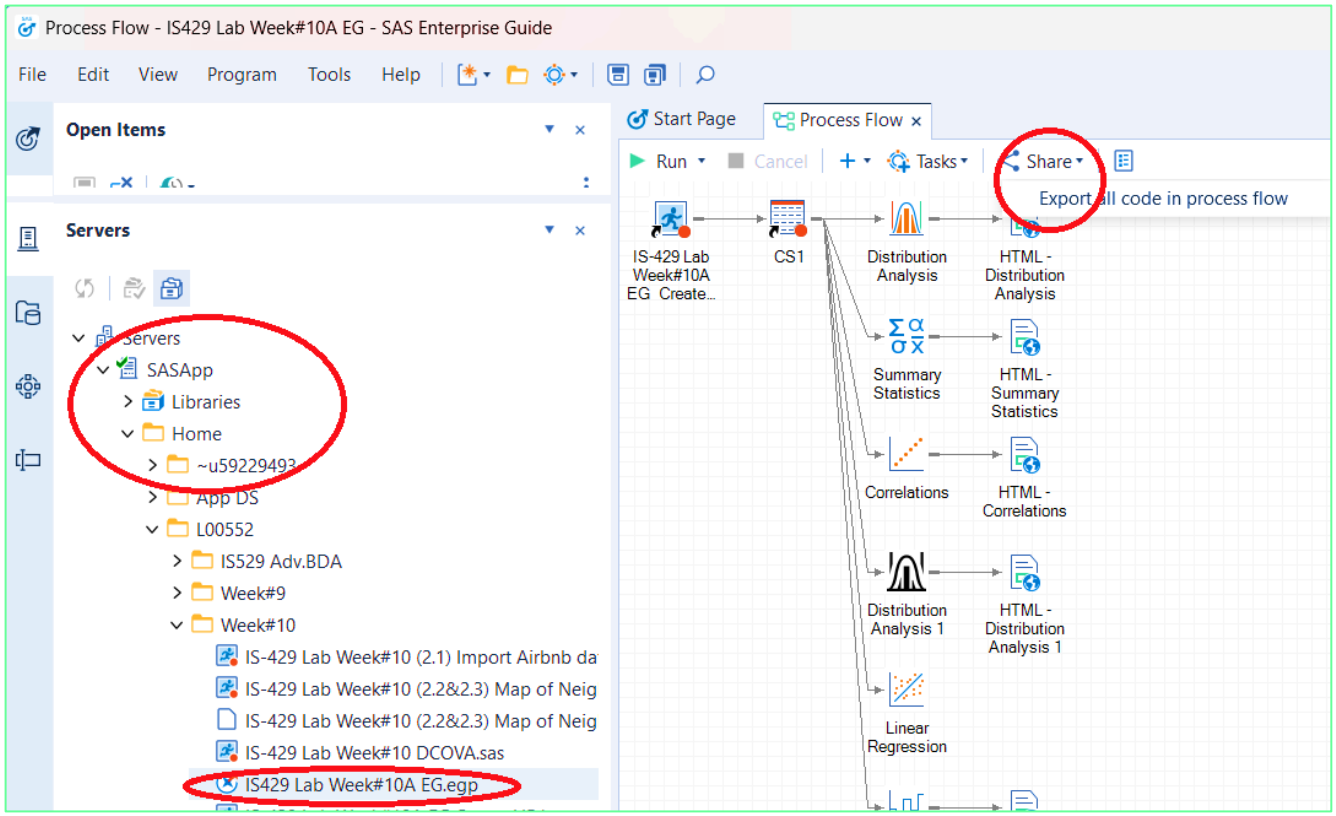
- SAS Enterprise Guide provides an intuitive project-based programming and point-and-click interface to SAS. It includes an intelligent program editor, querying capabilities, repeatable **process flows**, stored process creation and consumption, and a multitude of other features.

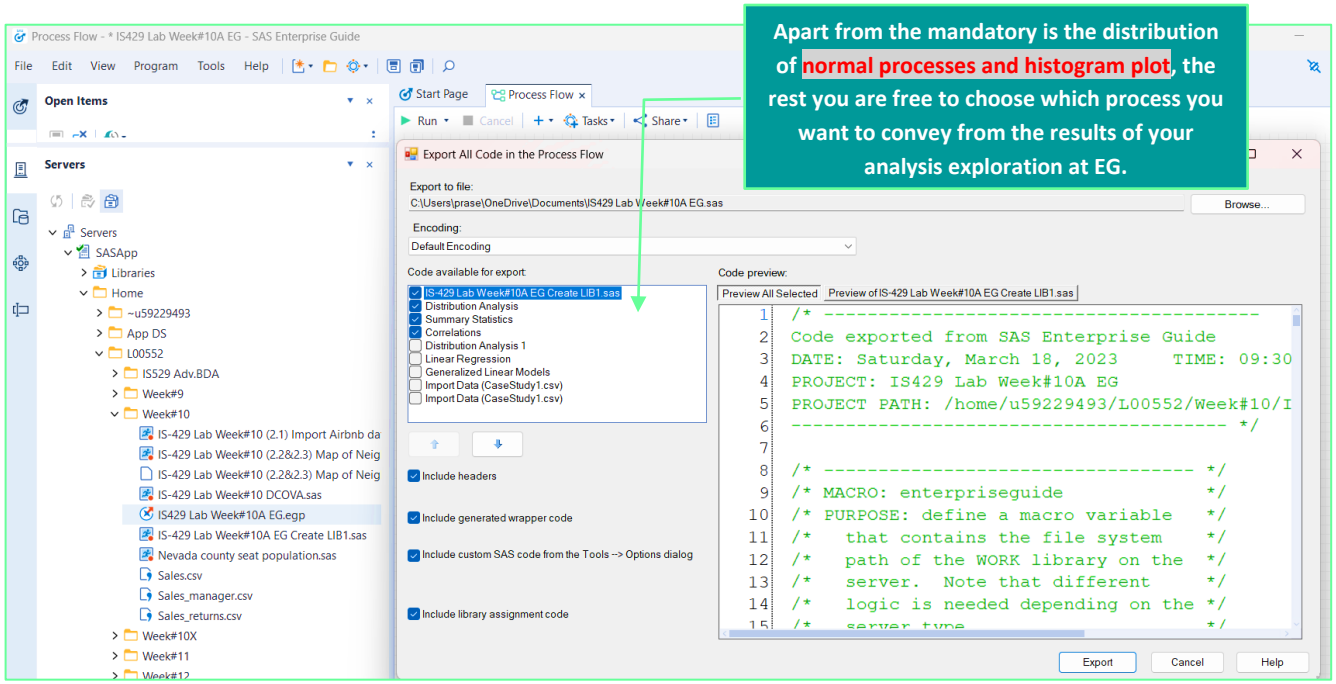
About Process Flows

- A process flow consists of one or more objects and could contain a process tree.
- An object is represented by an icon in the process flow.
- Examples of objects include a data set, a task, and your results. A process tree shows the relationship between two or more objects.
- SAS Enterprise Guide creates a default process flow when you start a project. For a single project, you can have multiple process flows.
- To view the process flow, **click Process Flow in the Project pane or the Process Flow tab in the work area**.

a. Screenshot of your process flows similar to the output **figure E** into the previous word file "**IS-429 Lab W#10 yourname (doc/docx)**".

b. Next save the EG Project that you created into the SAS Program (.SAS) using the Share feature as shown in the following figure:



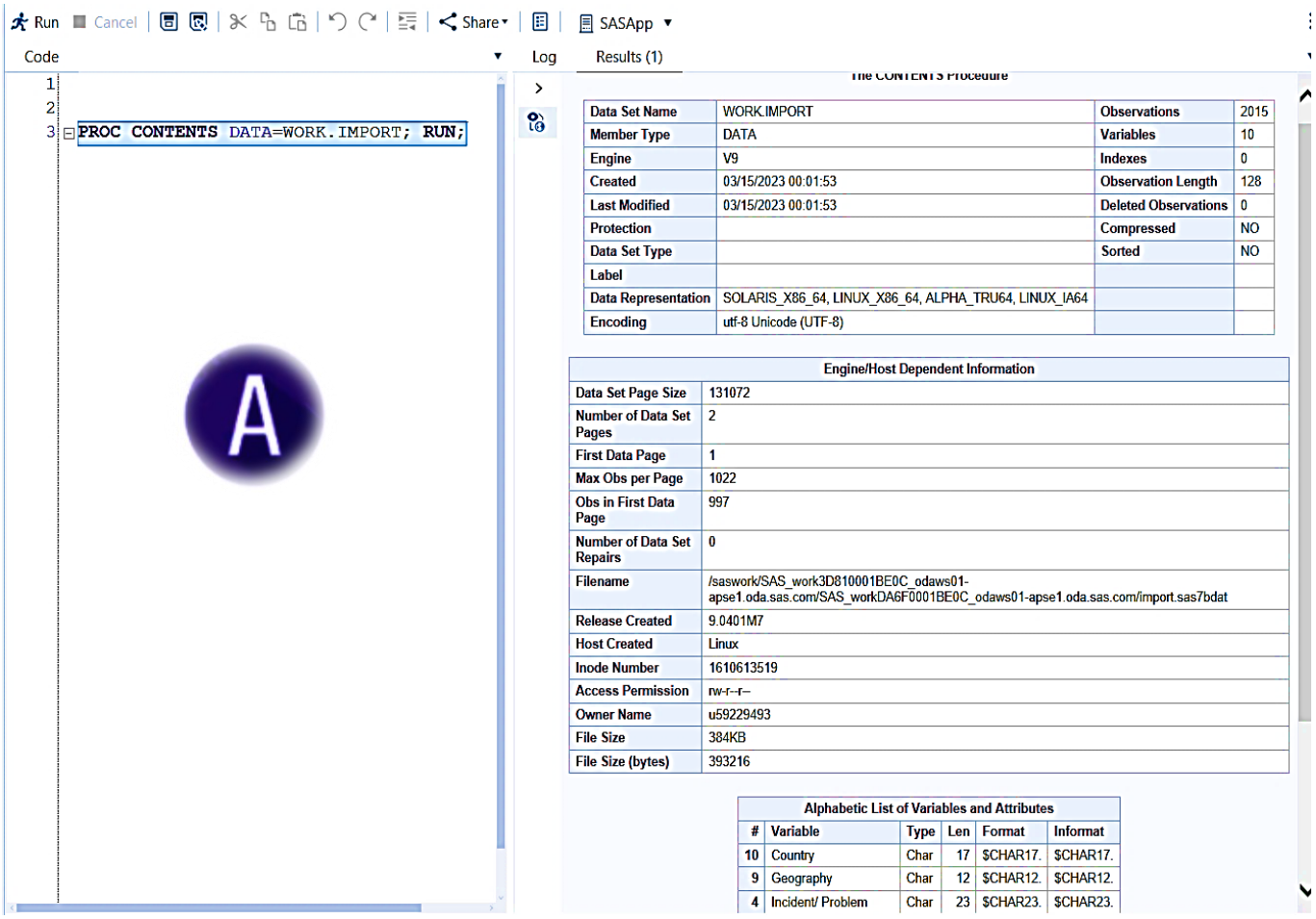


c. Name your EG project program into "IS-429 Lab W#10 EG yourNAME".

❖ The end of the Practicum Week#10 today is that you are asked to **collect Word files** containing screenshots and **SAS programs** from your EG project, in the form of **a zip file** format with the naming "**IS-429 Lab w#10 EG yourNAME.zip**" and immediately submit to the e-learning class of the Asynchronous and lab-synchronous.

RESULTS/ OUTPUT

A. Import the data into SAS cloud directory



The CONTENTS Procedure

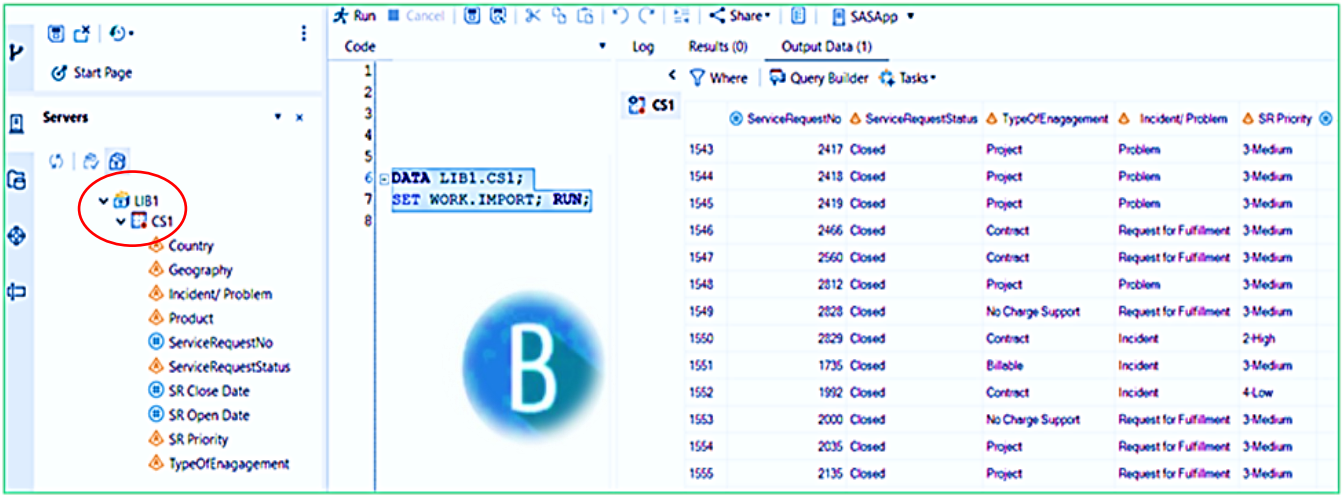
Data Set Name	WORK.IMPORT	Observations	2015
Member Type	DATA	Variables	10
Engine	V9	Indexes	0
Created	03/15/2023 00:01:53	Observation Length	128
Last Modified	03/15/2023 00:01:53	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	2
First Data Page	1
Max Obs per Page	1022
Obs in First Data Page	997
Number of Data Set Repairs	0
Filename	/saswork/SAS_work3D810001BE0C_odaws01-apse1.oda.sas.com/SAS_workDA6F0001BE0C_odaws01-apse1.oda.sas.com/import.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	1610613519
Access Permission	rw-r--r--
Owner Name	u59229493
File Size	384KB
File Size (bytes)	393216

Alphabetic List of Variables and Attributes

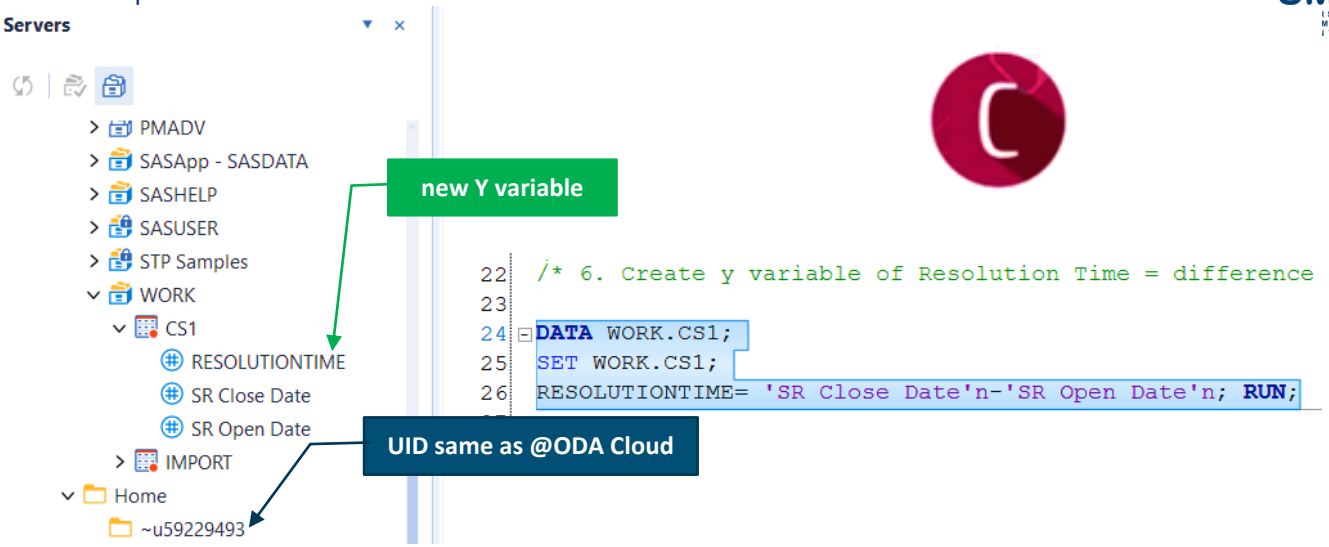
#	Variable	Type	Len	Format	Informat
10	Country	Char	17	$CHAR17.	$CHAR17.
9	Geography	Char	12	$CHAR12.	$CHAR12.
4	Incident/ Problem	Char	23	$CHAR23.	$CHAR23.

B. Store the SAS data set back in the SAS cloud library



	ServiceRequestNo	ServiceRequestStatus	TypeOfEnagagement	Incident/ Problem	SR Priority
1543	2417	Closed	Project	Problem	3-Medium
1544	2418	Closed	Project	Problem	3-Medium
1545	2419	Closed	Project	Problem	3-Medium
1546	2466	Closed	Contract	Request for Fulfillment	3-Medium
1547	2560	Closed	Contract	Request for Fulfillment	3-Medium
1548	2812	Closed	Project	Problem	3-Medium
1549	2828	Closed	No Charge Support	Request for Fulfillment	3-Medium
1550	2829	Closed	Contract	Incident	2-High
1551	1735	Closed	Billable	Incident	3-Medium
1552	1992	Closed	Contract	Incident	4-Low
1553	2000	Closed	No Charge Support	Request for Fulfillment	3-Medium
1554	2035	Closed	Project	Request for Fulfillment	3-Medium
1555	2135	Closed	Project	Request for Fulfillment	3-Medium

C. Define the problem: Clean the data to resolve the SR



D. Analyze, Visualize, and get Insights: Understand the distribution of the y variable

Distribution analysis of: RESOLUTIONTIME

The UNIVARIATE Procedure
Variable: RESOLUTIONTIME

Basic Statistical Measures			
Location		Variability	
Mean	56.42044	Std Deviation	71.08331
Median	28.00000	Variance	5053
Mode	0.00000	Range	484.00000
		Interquartile Range	74.00000

Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	56.42044	53.12615	59.71472
Std Deviation	71.08331	68.82936	73.49099
Variance	5053	4737	5401

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	33.59051	Pr > |t|	<.0001
Sign	M	809.5	Pr >= |M|	<.0001
Signed Rank	S	655695	Pr >= |S|	<.0001

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	224	11.12	100.00

Generated by SAS ('SASApp', Linux) on March 15, 2023 at 01:05:37 AM

Distribution analysis of: RESOLUTIONTIME

The UNIVARIATE Procedure
Fitted Normal Distribution for RESOLUTIONTIME

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	56.42044
Std Dev	Sigma	71.08331

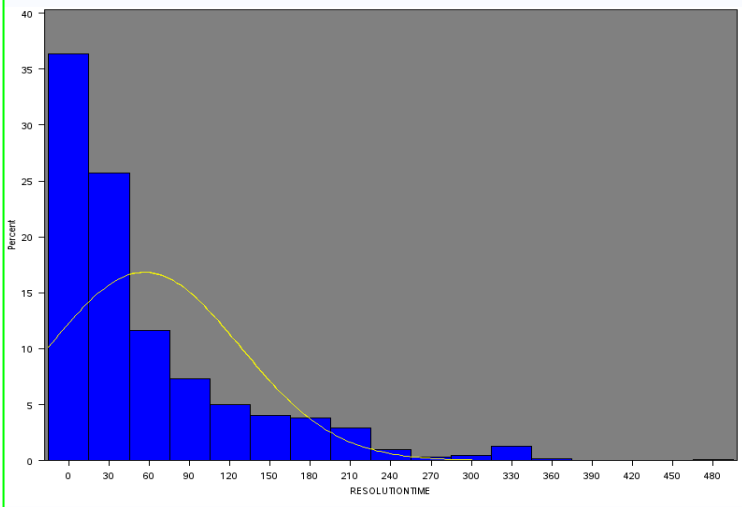
Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.213678	Pr > D	<0.010
Cramer-von Mises	W-Sq	24.550534	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	134.845013	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.000	-108.94407
5.0	0.000	-60.50121
10.0	1.000	-34.67649
25.0	6.000	8.47547
50.0	28.000	56.42044
75.0	80.000	104.36540
90.0	161.000	147.51736
95.0	206.000	173.34208
99.0	320.000	221.78494

Generated by SAS ('SASApp', Linux) on March 15, 2023 at 01:05:37 AM

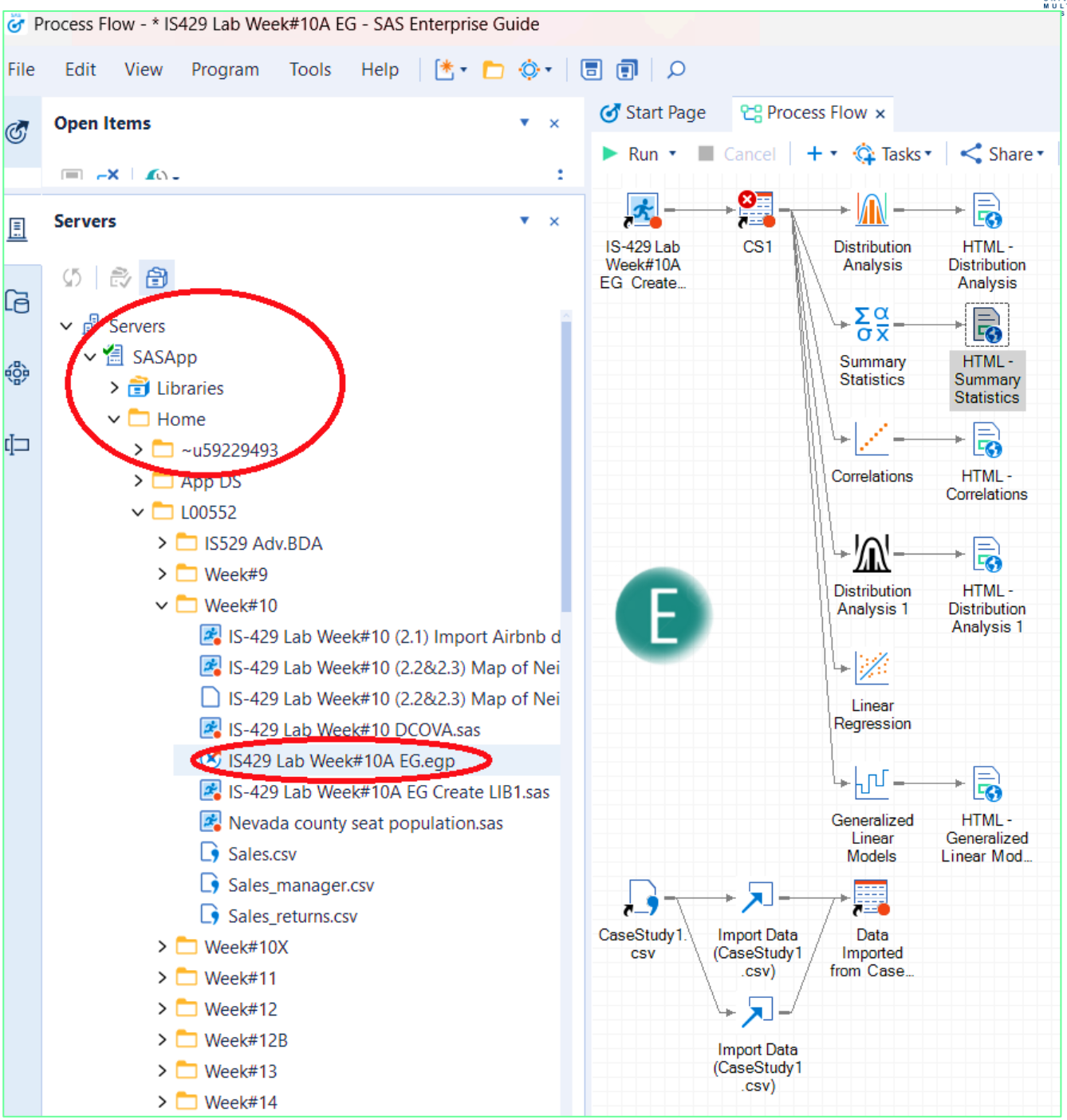
Distribution analysis of: RESOLUTIONTIME

The UNIVARIATE Procedure



Generated by SAS ('SASApp', Linux) on March 15, 2023 at 01:05:37 AM

E. Process Flows



REFERENCE

1. Subhashini Sharma Tripathi. 2016. Learn Business Analytics in Six Steps Using SAS® and R. Apress®.
2. SAS® Support | Documentation
3. Other additional references are excerpts from various Online Learning/websites.

The End