

VISVESVARAYA TECHNOLOGICAL UNIVERSITY
“Jnana Sangama”, Belgaum 590014, Karnataka, India.



Blue Prism Project Report
On
“WEB SCRAPING”

Submitted in Partially fulfillment of the requirement for the award of degree

of

Bachelor of Engineering

In

Computer Science & Engineering

Of Visvesvaraya Technological University, Belgaum.

Submitted by:

AMAN HARIS (1AM18CS016)

Project Guide:

Ms. Sonali S Dash

Asst Professor, Dept. of CS

Ms. Sowmya BK

Asst Professor, Dept. of CS



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
AMC ENGINEERING COLLEGE**

(NAAC & NBA Accredited, Approved by AICTE, New Delhi & Affiliated to VTU, Belagavi)



18th K.M. Bannerghatta Main Road, Bengaluru – 560083



(Accredited by NAAC & NBA, MINISTRY OF HRD, NEW DELHI & Affiliated to VTU Belagavi)

Year 2021-22

AMC ENGINEERING COLLEGE

(NAAC & NBA Accredited, Approved by AICTE, New Delhi & Affiliated to VTU,
Belagavi, Bengaluru- 560083)

Department of Computer Science & Engineering



CERTIFICATE

Certified that the Project work entitled "**WEB SCRAPING**" carried out by bona fide Students **AMAN HARIS (1AM18CS016)** of AMC Engineering College, in partial fulfillment for the award of Bachelor of Engineering in Computer Science & Engineering of Visvesvaraya Technological University, Belgaum during the year 2020-2021. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of Project Work prescribed for the said Bachelor of Engineering degree.

Project Guide

Ms. Sonali S Dash

Ms. Sowmya B K

Dept. of CSE

HOD

Dr. Latha C A

Head of Dept. CSE, AMCEC

Principal

Dr. A G Nataraj

Principal, AMCEC

External Viva

External Name

1. _____

2. _____

Signature with Date

1. _____

2. _____

DECLARATION

We the undersigned students of 6th semester Department of Computer Science & Engineering, AMC Engineering College, declare that our project work entitled “Web Scraping**” is a bona fide work of ours. Our project is neither a copy nor by means a modification of any other engineering project.**

We also declare that this project was not entitled for submission to any other university in the past and shall remain the only submission made and will not be submitted by us to any other university in the future.

NAME

USN

SIGNATURE

AMAN HARIS

1AM18CS016

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of this project report would be complete only with mention of people who made it possible, whose support rewarded our effort with success.

First and foremost, I would like to thank the almighty.

I consider it my privilege to express my deep gratitude to our Chairman Dr. K R Paramahamsa for providing the infrastructure and an opportunity to pursue my degree in his prestigious institution

I express my heartfelt gratitude and humble thanks to Dr. Latha C A Prof., & HOD Department of Computer Science & Engineering, for the facilities and support during our project.

I greatly express my gratitude to my guide Ms. Sonali S Dash and Ms. Sowmya BK, Associate Professor, Dept., of Computer Science & Engineering, for taking a keen interest in the progress of the project suggestions given to me.

I also would like to acknowledge all the teaching and non-teaching staff of the department for their kind cooperation during the project.

In the end, I sincerely thank my beloved family members and friends for their valuable suggestions, encouragement and unwavering support.

AMAN HARIS

(1AM18CS016)

ABSTRACT

Robotic process automation (RPA) is the application of technology that provides organizations with a digital workforce that follows rule-based business processes and interacts with the organizations systems in the same way that existing users currently do. Blue Prism is one of the RPA Tool has been credited for coining the term Robotic Process Automation.

The aim of the project is to collect data such as name, prize and rating of an item from an e-commerce websites and store it into a collection so that we can get different costs available for searched item. With the merge option we group together different data into one collection which is then sorted to find out the least cost for the searched item. This data is then written on to an excel sheet.

Here we used the major parts of Blue Prism's Process Studio and Object Studio And inbuilt VBO like Utility Collection Manipulation and MS Excel to carry out the automated process.

CONTENTS

CHAPTER NAME	PAGE NO.
1. Introduction.....	7
1.1 Introduction to Blue Prism	
2. System Requirement.....	8
2.1 Hardware requirements	
2.2 Software requirements	
3. System Architecture.....	9
4. Implementation	12
5. Snapshots	14
6. Conclusion	27
7. Future Enhancement	28
8. References	29

LIST OF FIGURES:

<u>FIG NO.</u>	<u>TITLE</u>	<u>PAGE NO.</u>
Fig 1.1	Design	9
Fig 1.2	Data Items	10
Fig 1.3	Design of Comparison data	11

INTRODUCTION

1.1 INTRODUCTION TO BLUE PRISM

Blue prism is a UK-Based Software Company and is one of the leading robotic process automation tools. It is used to automate mundane tasks such that they could operate without any manual intervention. Blue prism has gained edge over its competitors as it has better security, flexibility, scalability, compliance, and resilience. Blue Prism helps the enterprises to automate the business operations in an agile and cost-effective manner.

BLUE PRISMS ROBOTIC AUTOMATION Robotic Automation refers to process automations where computer software drives existing enterprise application software in the same way that a user does. This means unlike traditional application software, Robotic Automation is a tool or platform that operates and orchestrates other application software through the existing application's user interface and in this sense is not "integrated".

Blue Prism's Robotic Automation software enables business operations to be agile and cost-effective through rapid automation of manual, rule-based, back office administrative processes, reducing cost and improving accuracy by creating a "virtual workforce".

The virtual workforce is built by the operational teams or accredited Blue Prism partners using our robotic automation technology to rapidly build and deploy automations through leveraging the presentation layer of existing enterprise applications. The automations are configured and managed within an IT-governed framework and operating model which has been iteratively developed through numerous large scale and complex deployments.

CHAPTER 2

SYSTEM REQUIREMENTS

2.1 SOFTWARE REQUIREMENTS:

- OPERATING SYSTEM : Windows 8, Windows 10
- FRONT END : Blue Prism
- BACK END : Internet explorer

2.2 HARDWARE REQUIREMENT:

- SYSTEM : Pentium IV 2.4 GHz or above
- HARD DISK : 40 GB, 80 GB, 160 GB or above
- MONITOR : 15 VGA color
- RAM : 4 GB or above

System Architecture

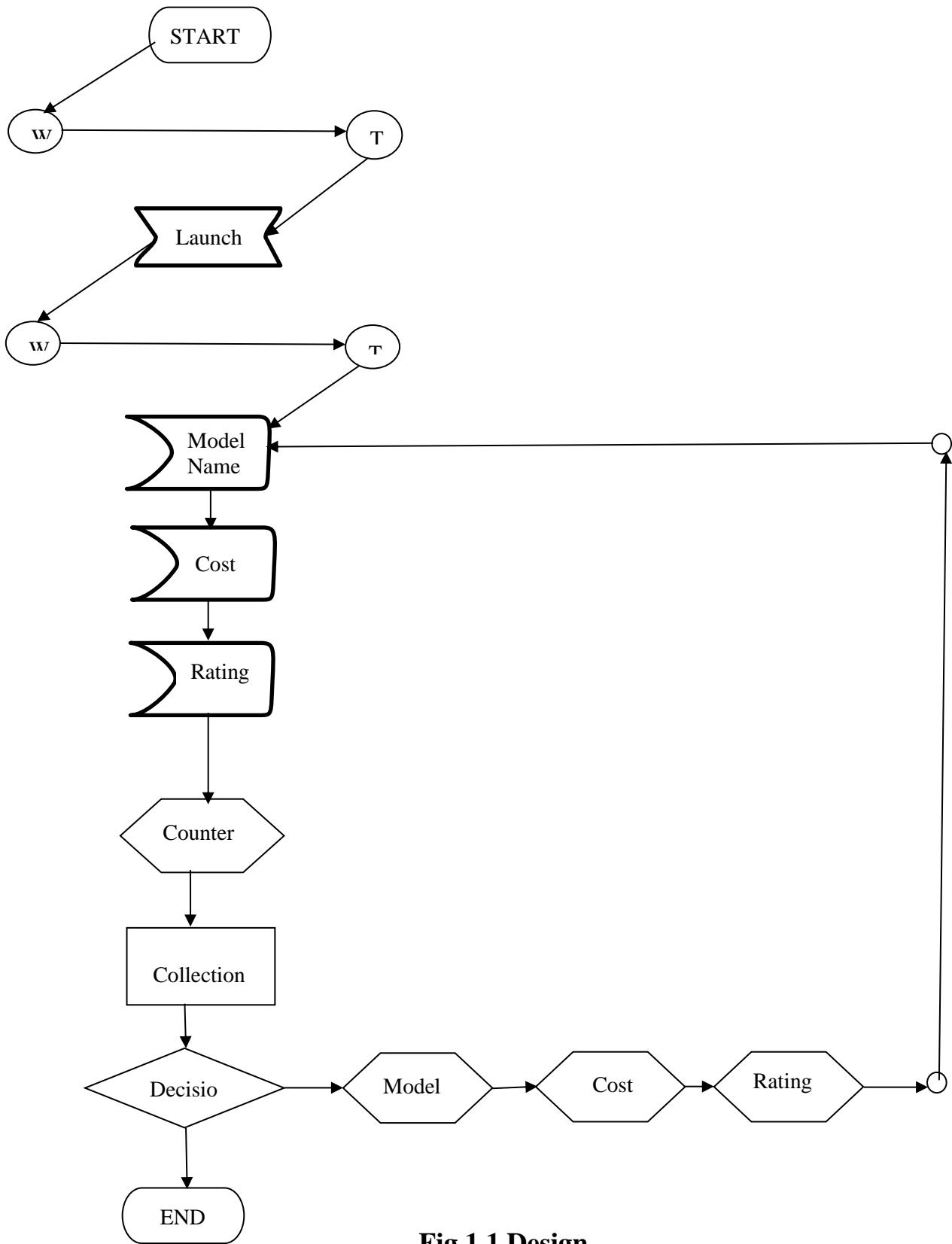


Fig 1.1 Design

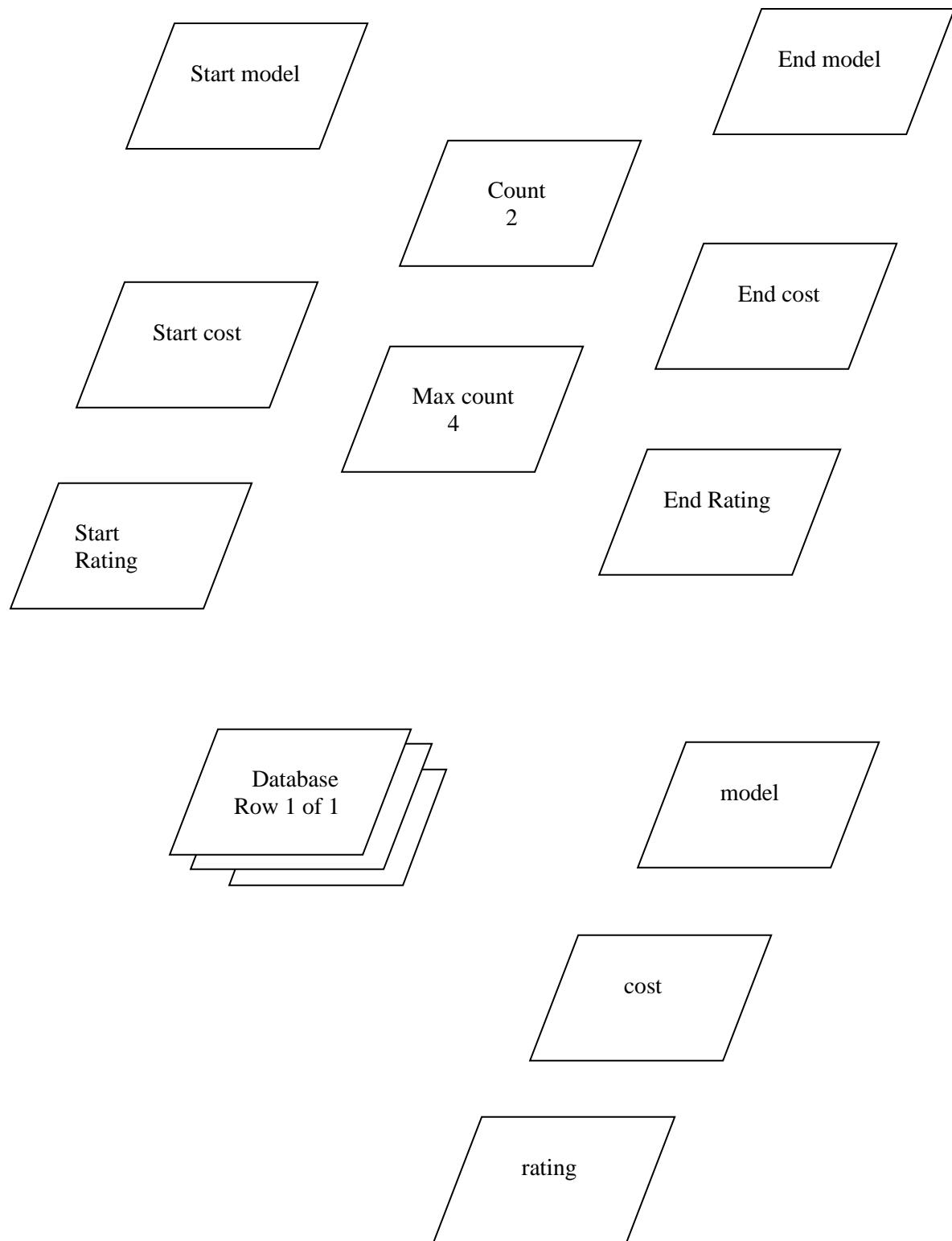


Fig 1.2 Data Items

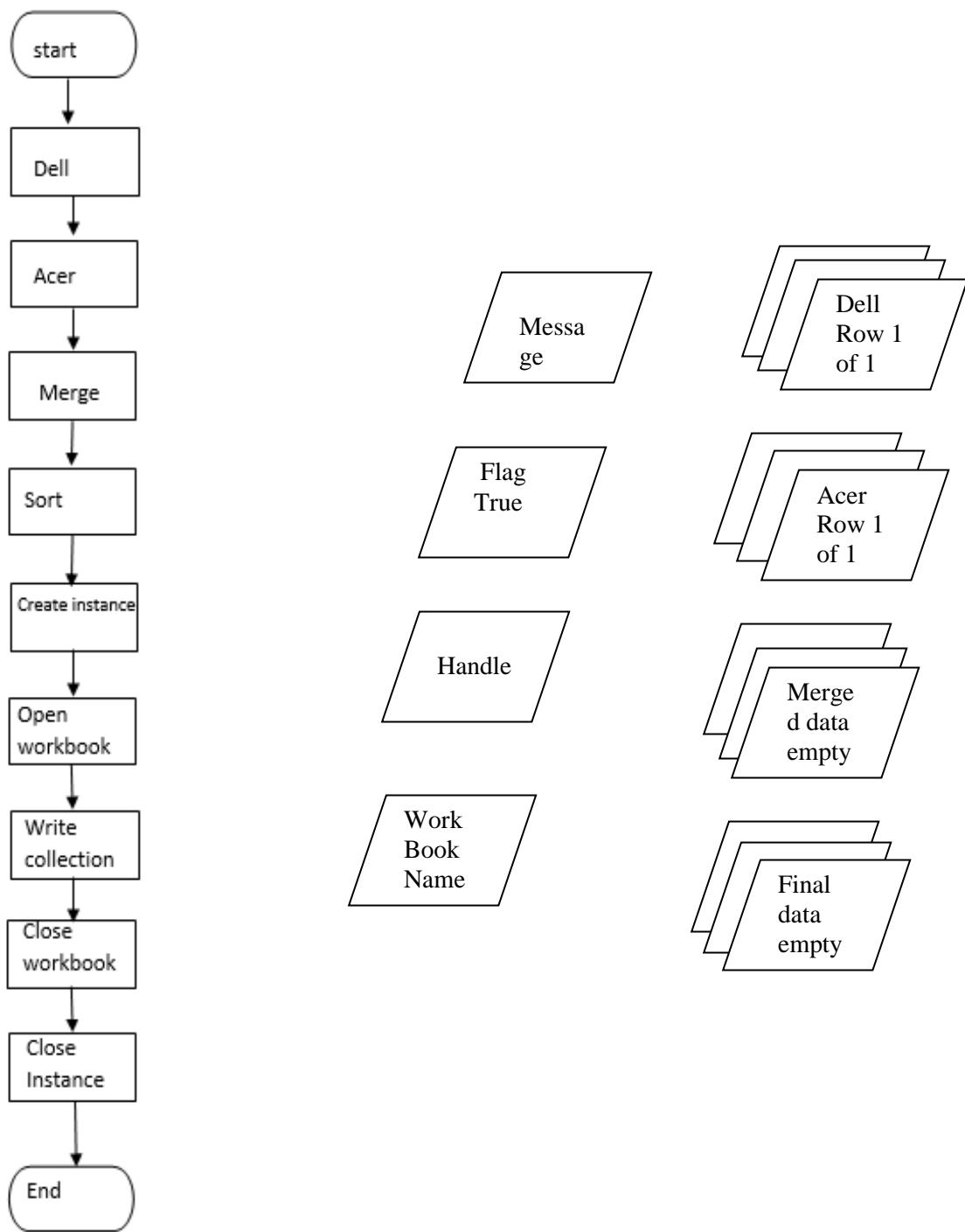
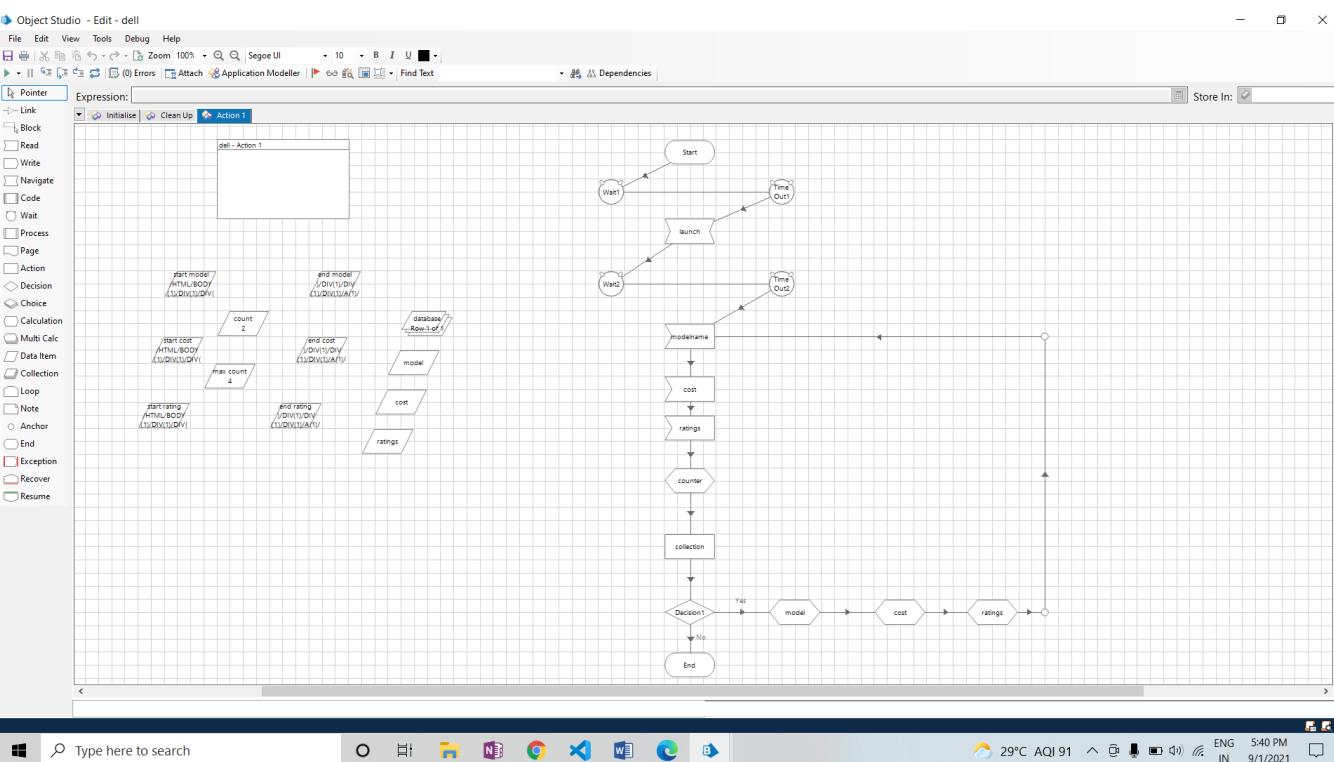
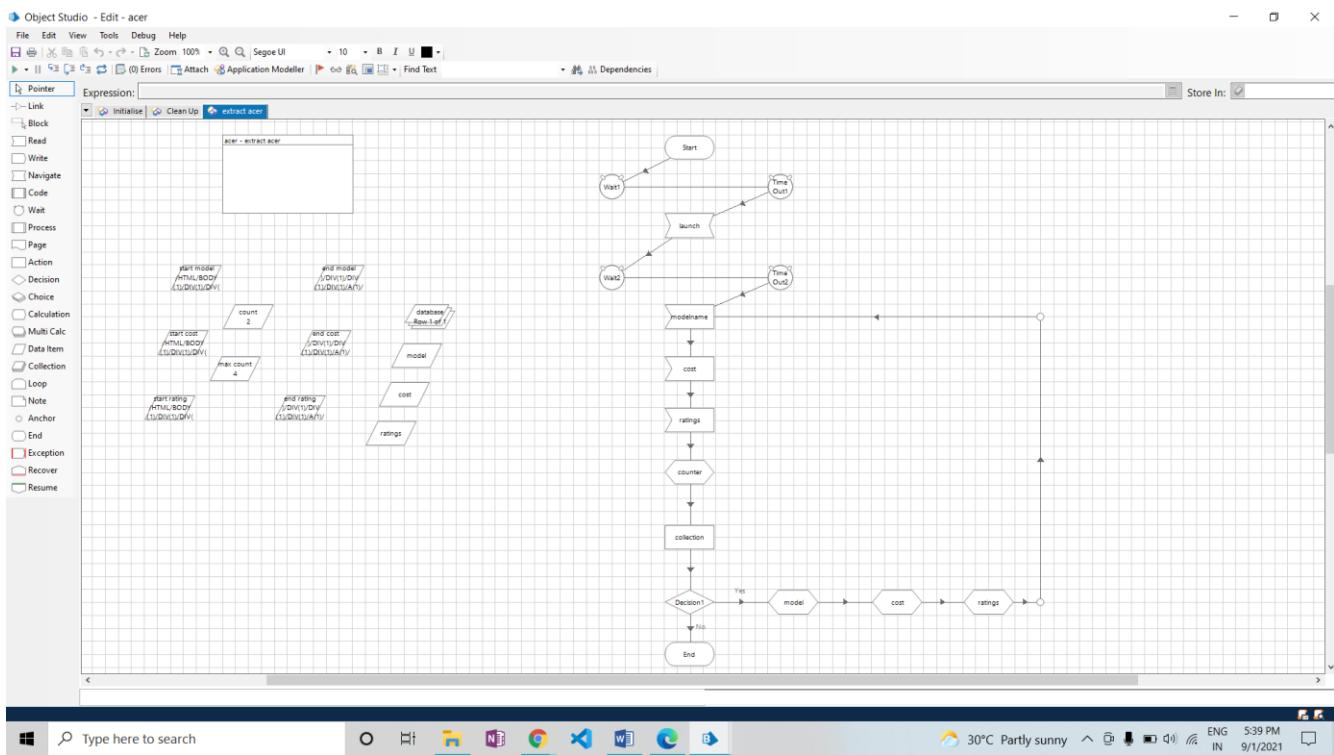
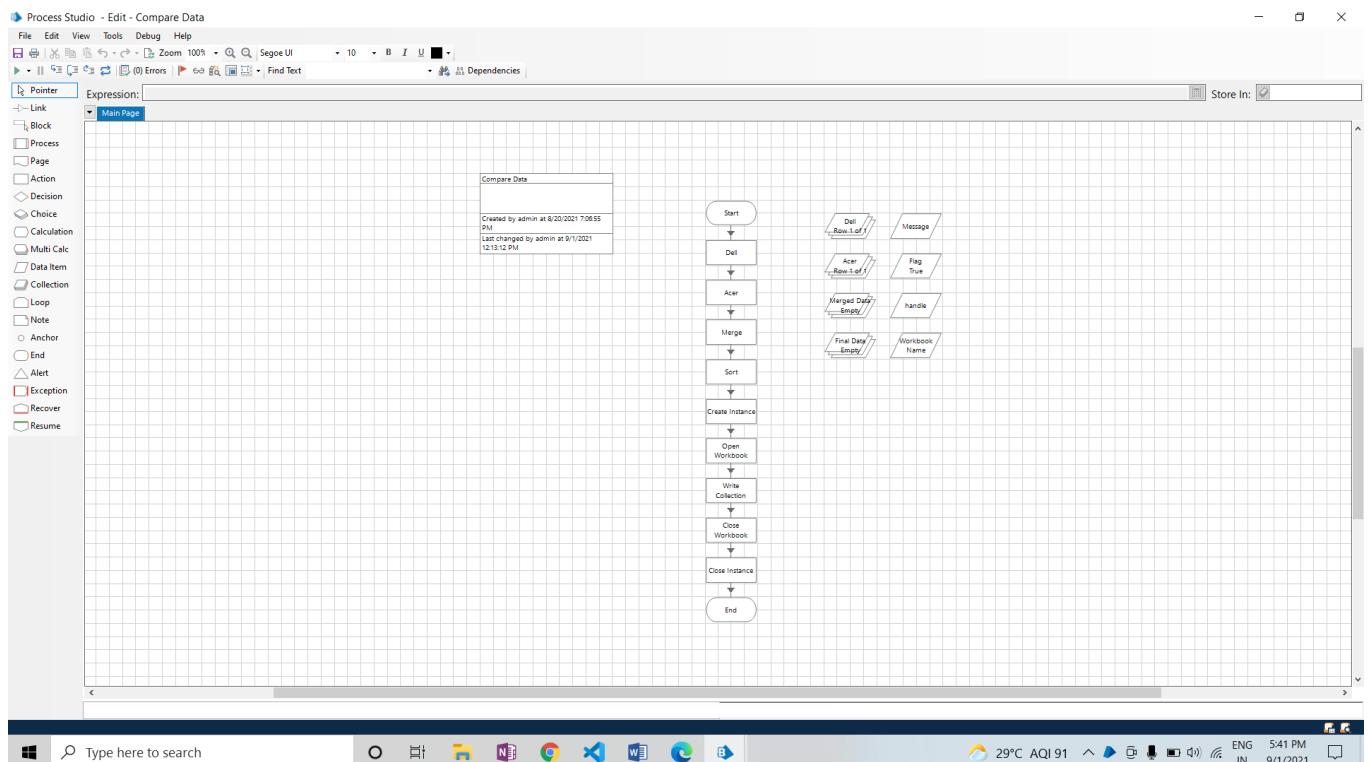


Fig 1.3
Design of
Comparison Data

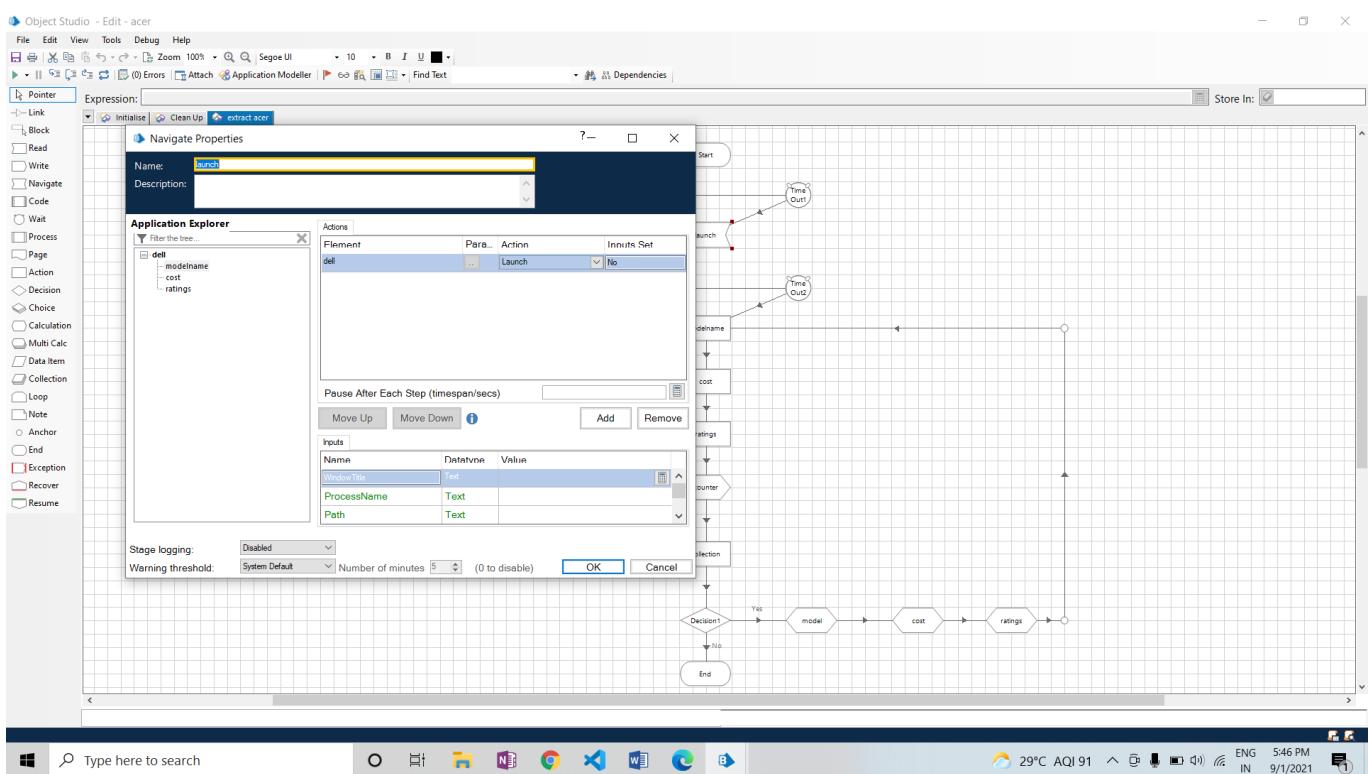
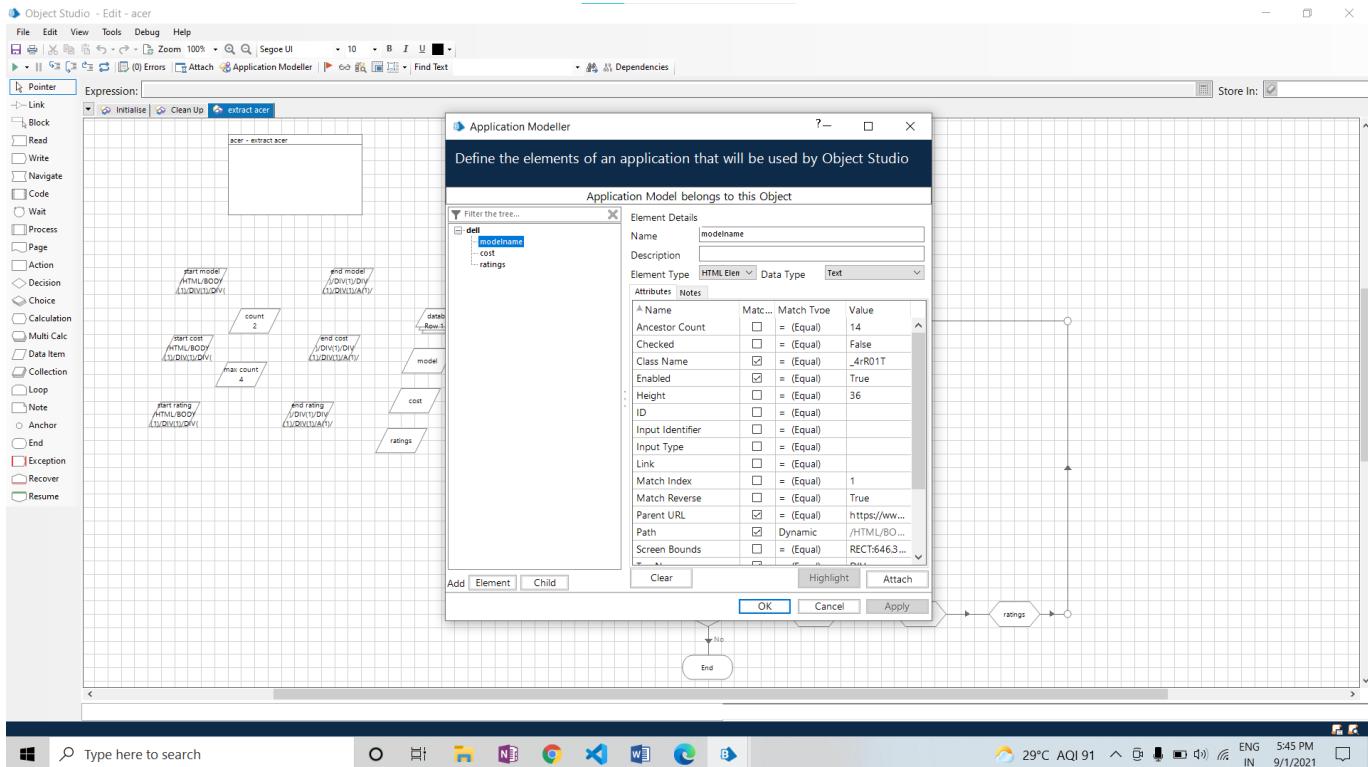
Implementation

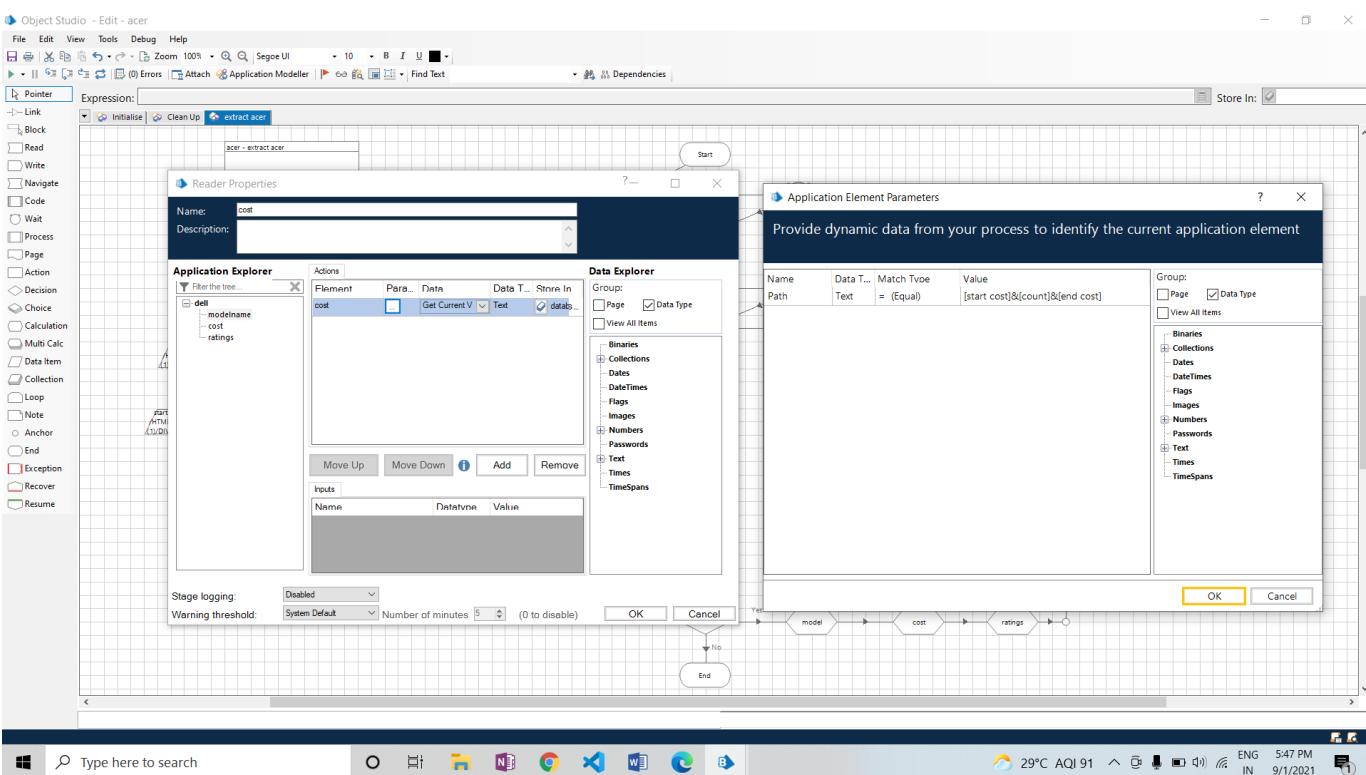
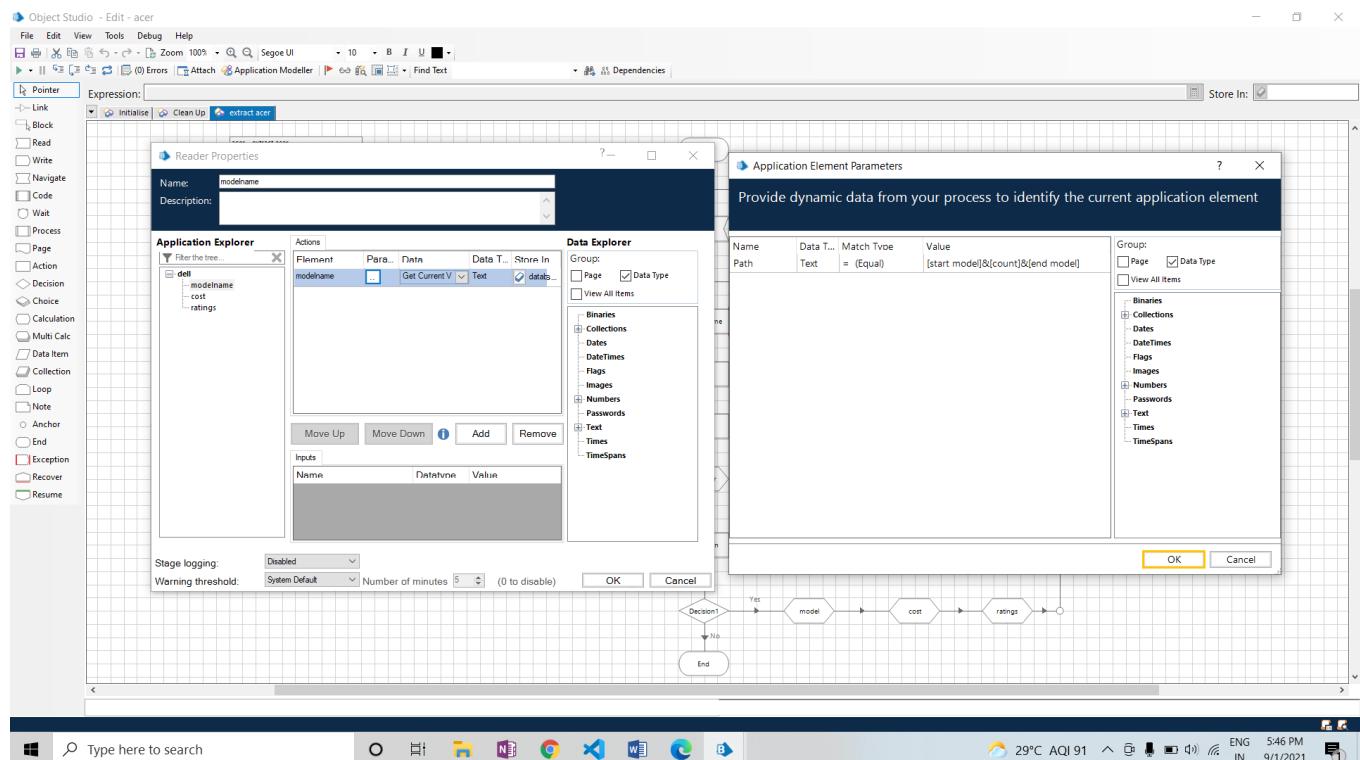


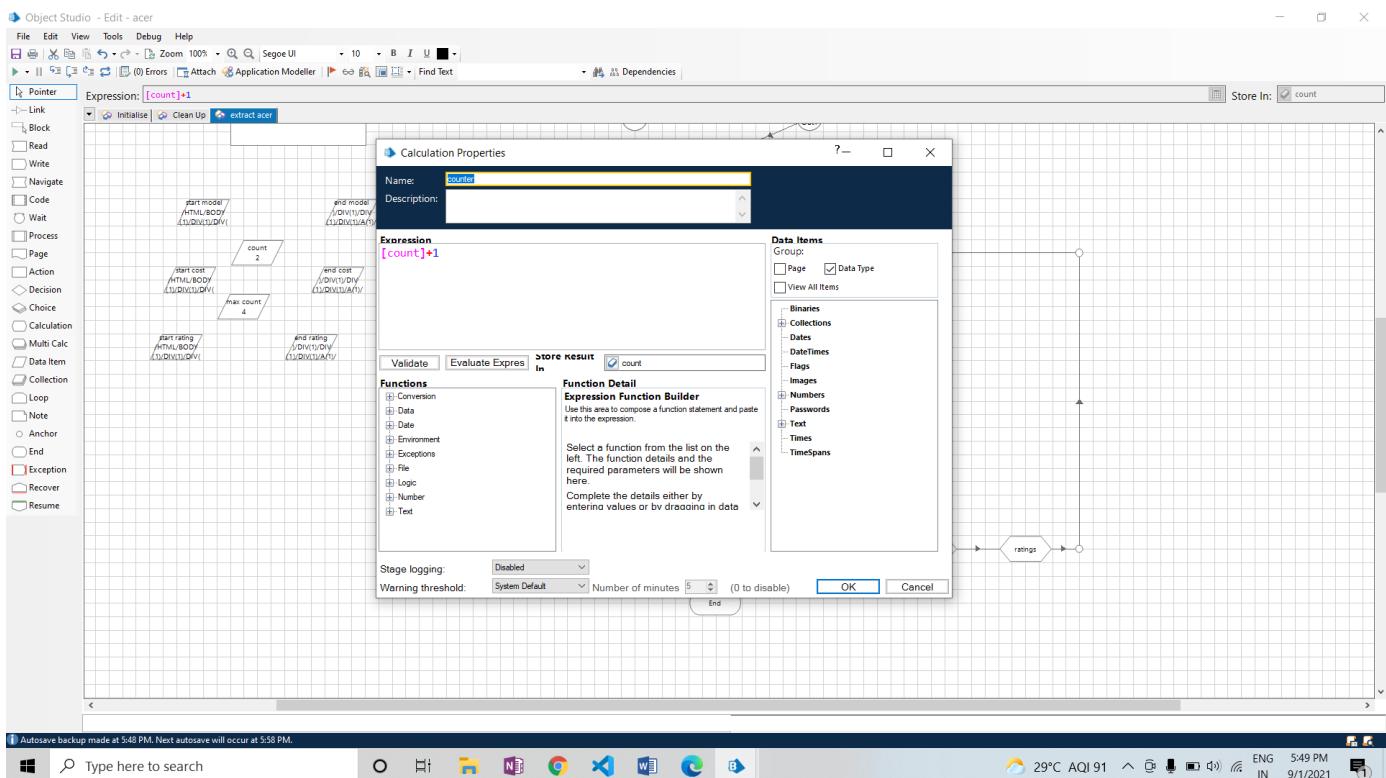
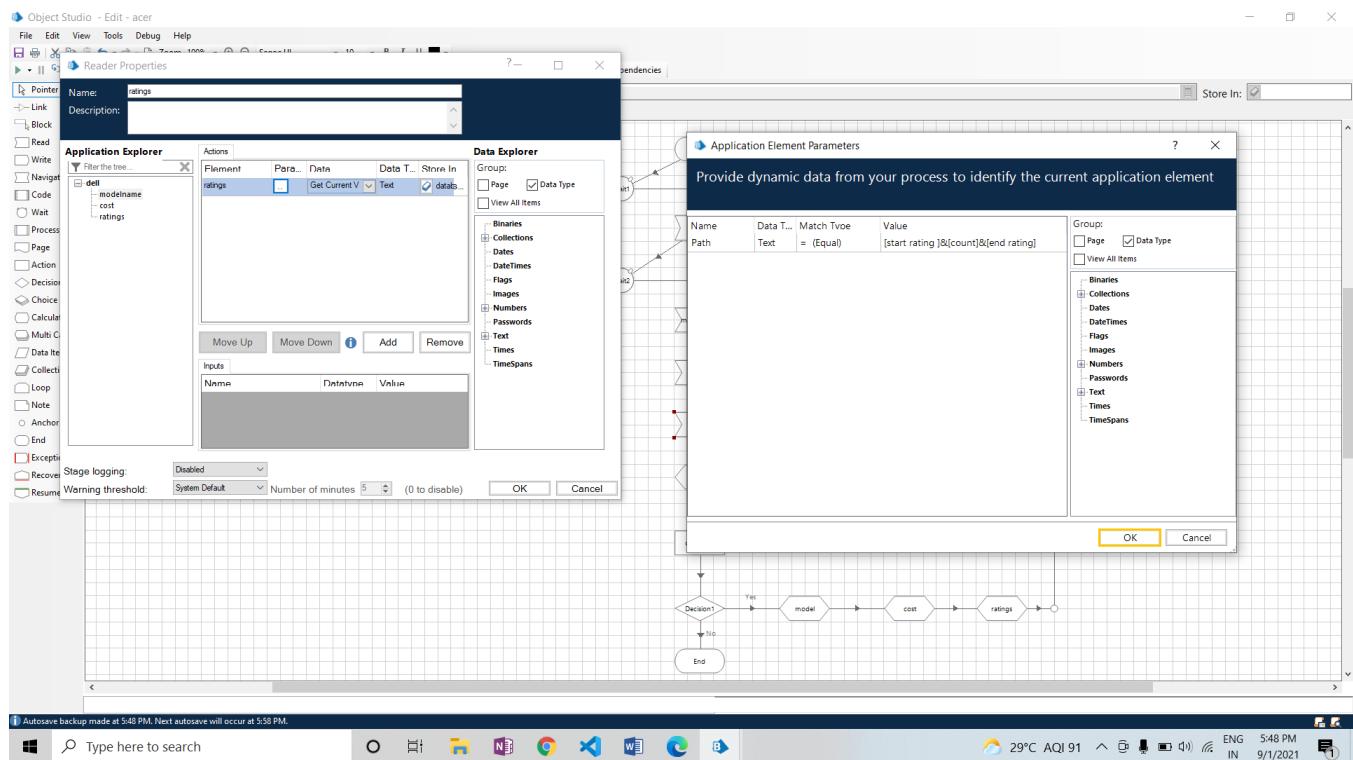


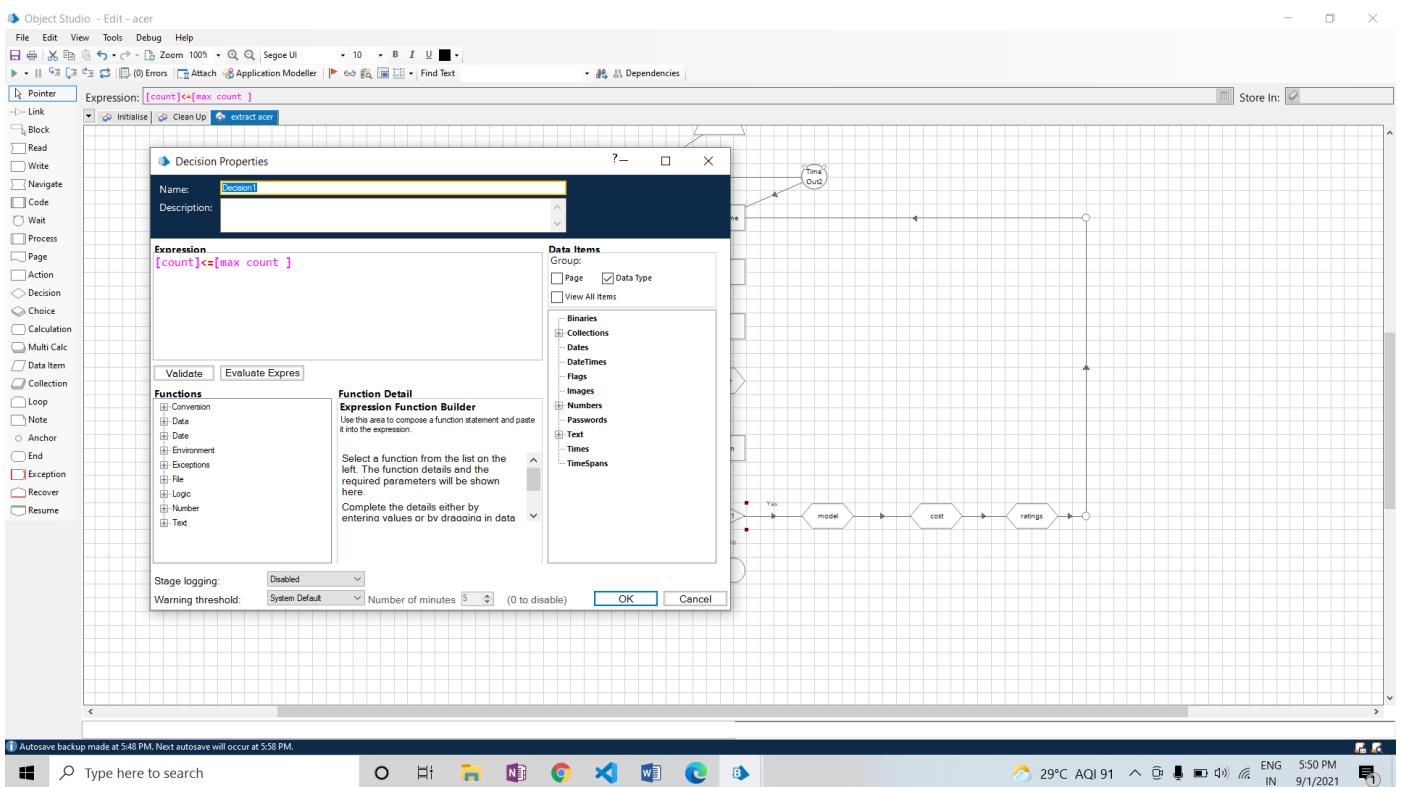
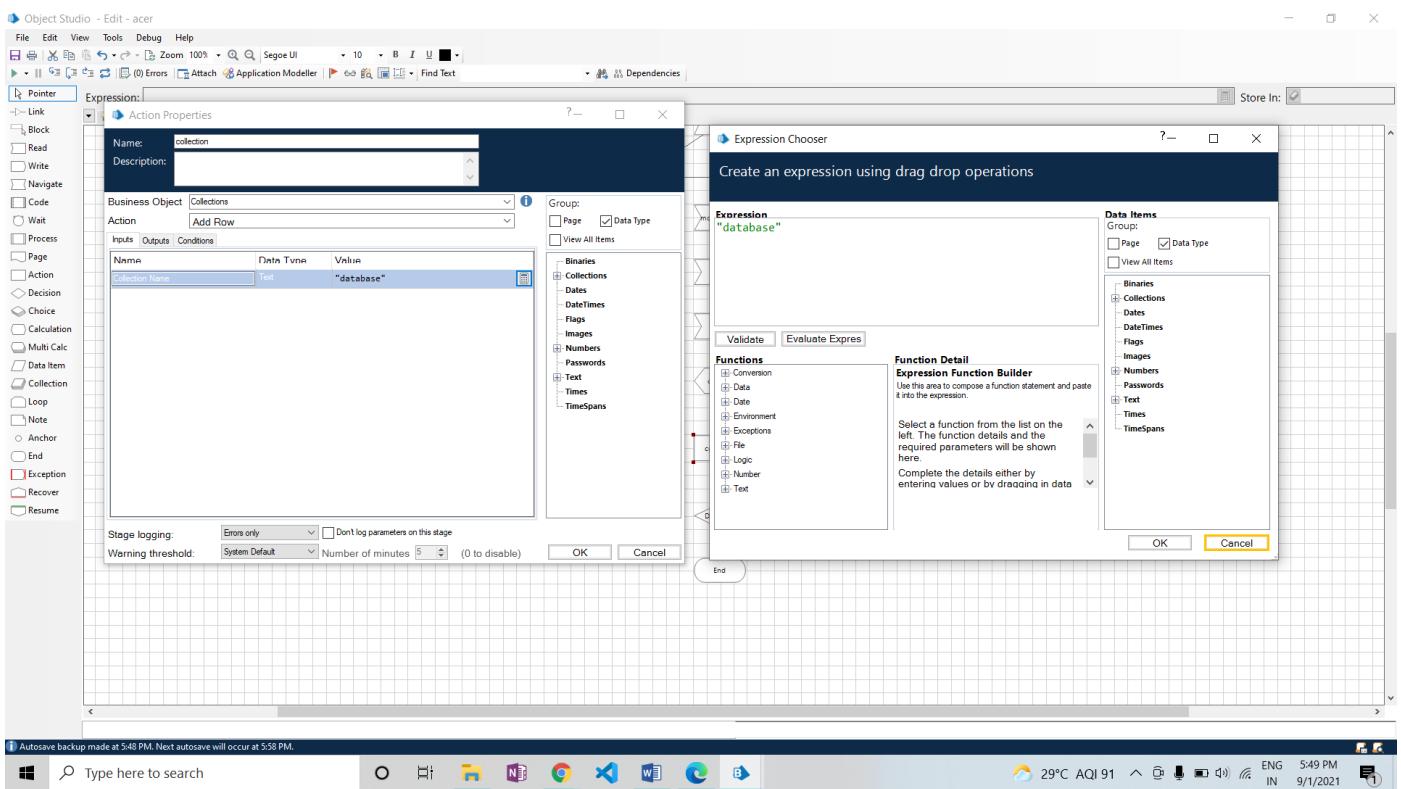
CHAPTER 5

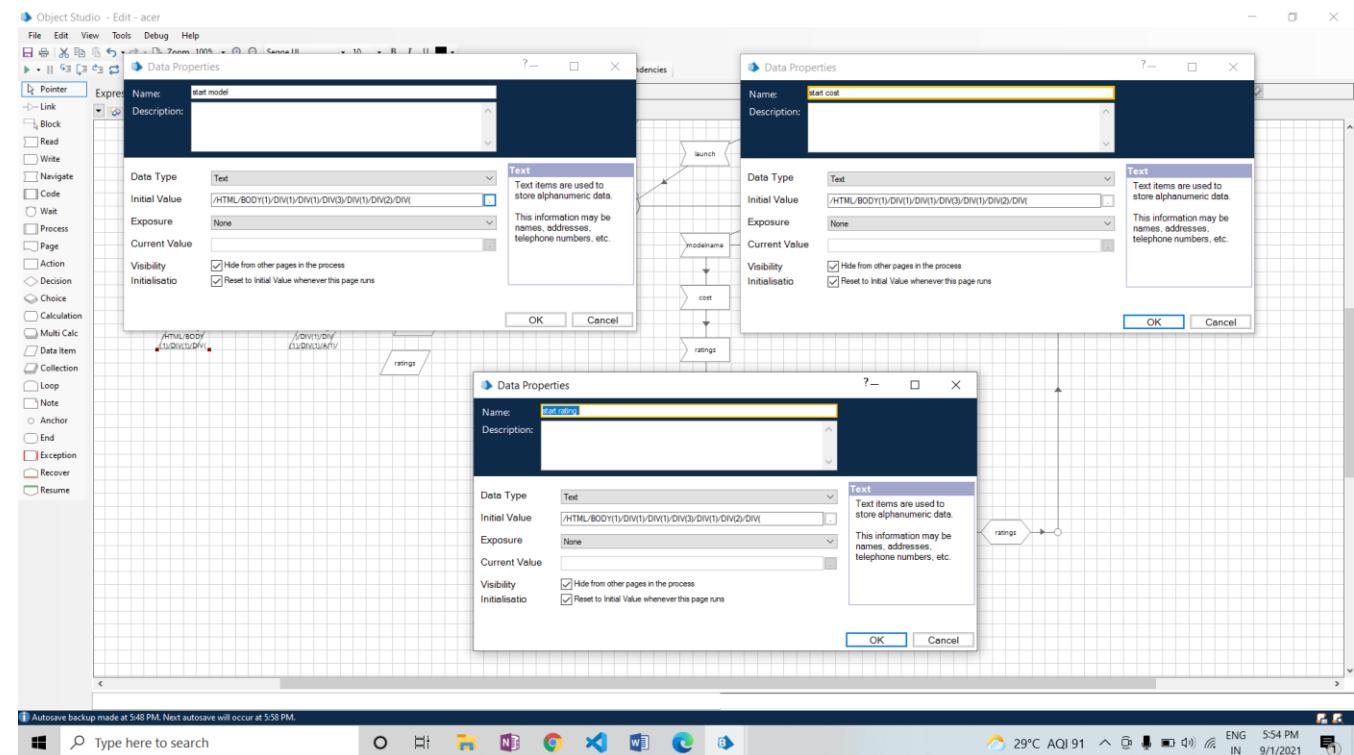
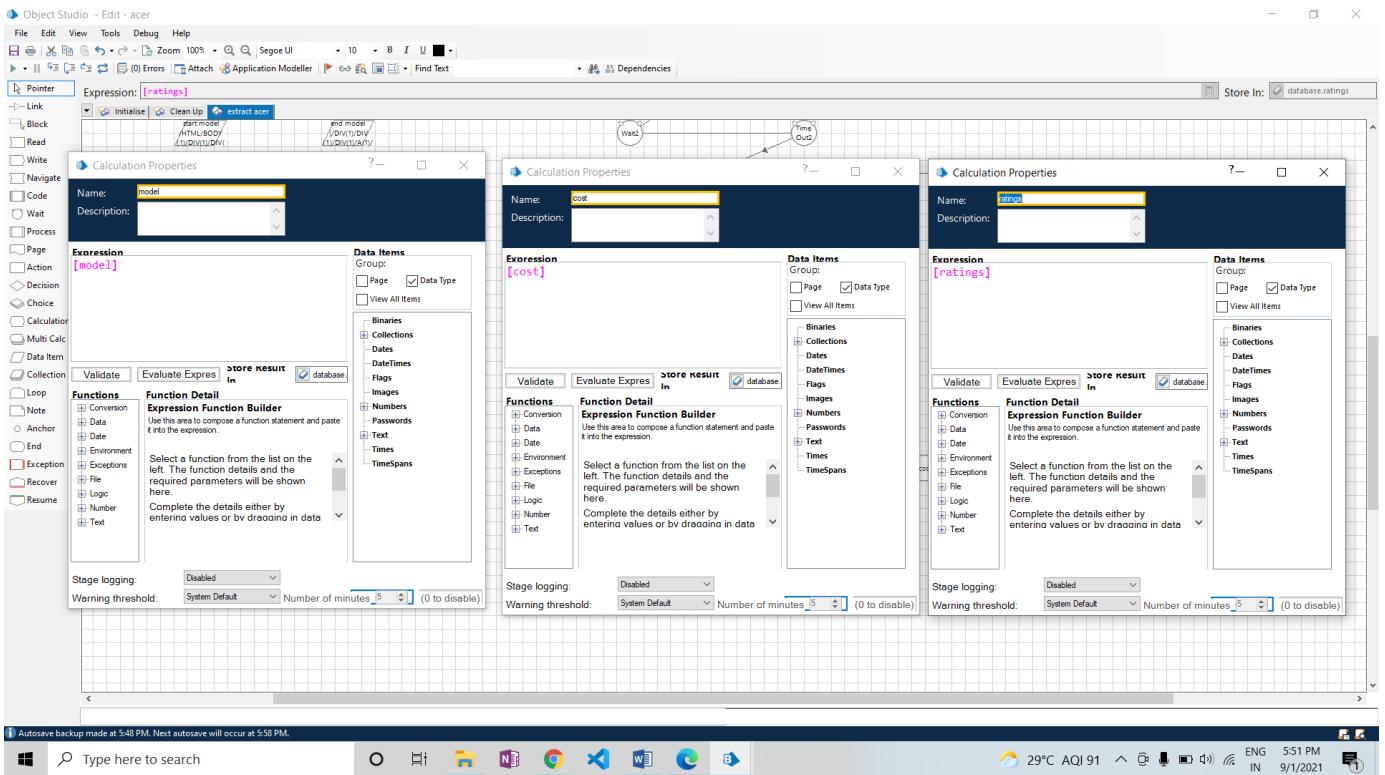
SNAPSHOTS

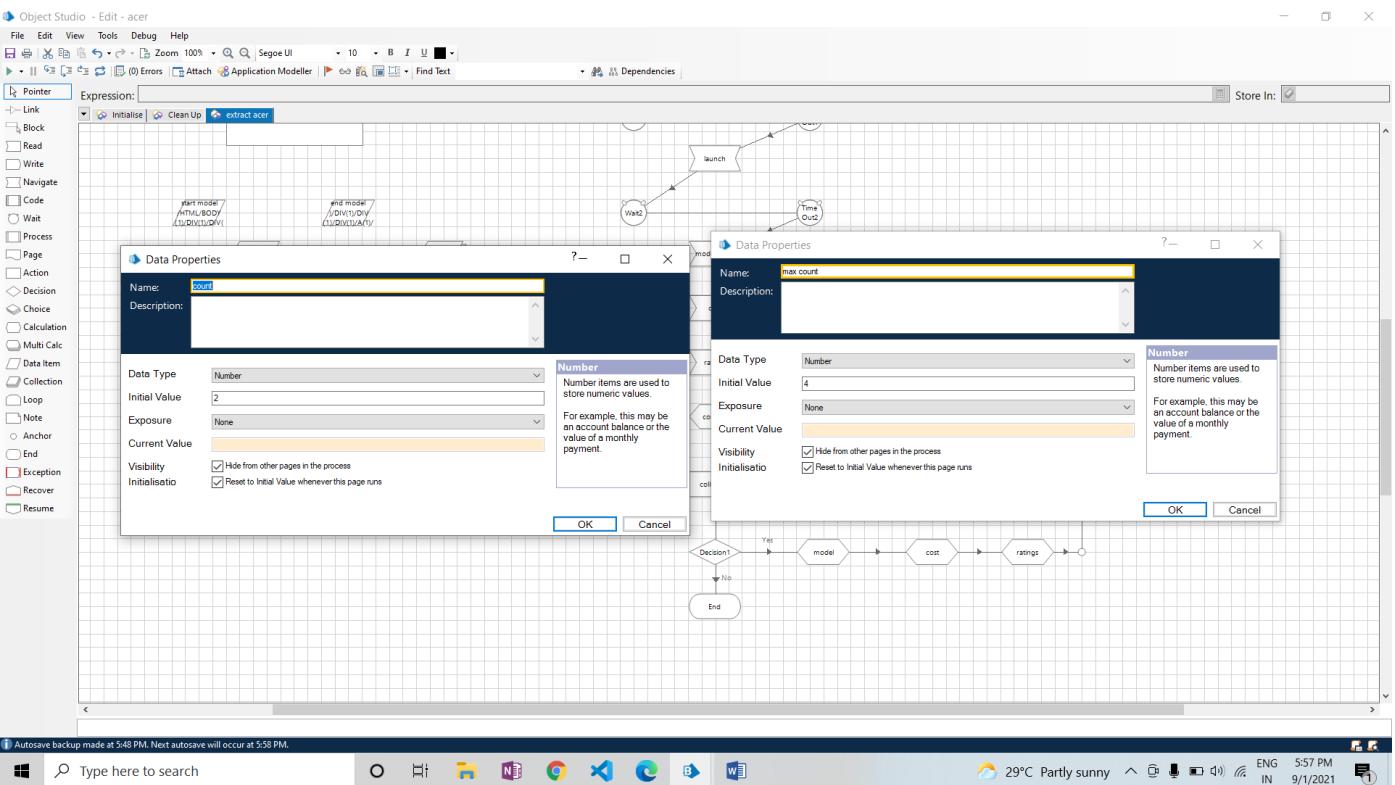
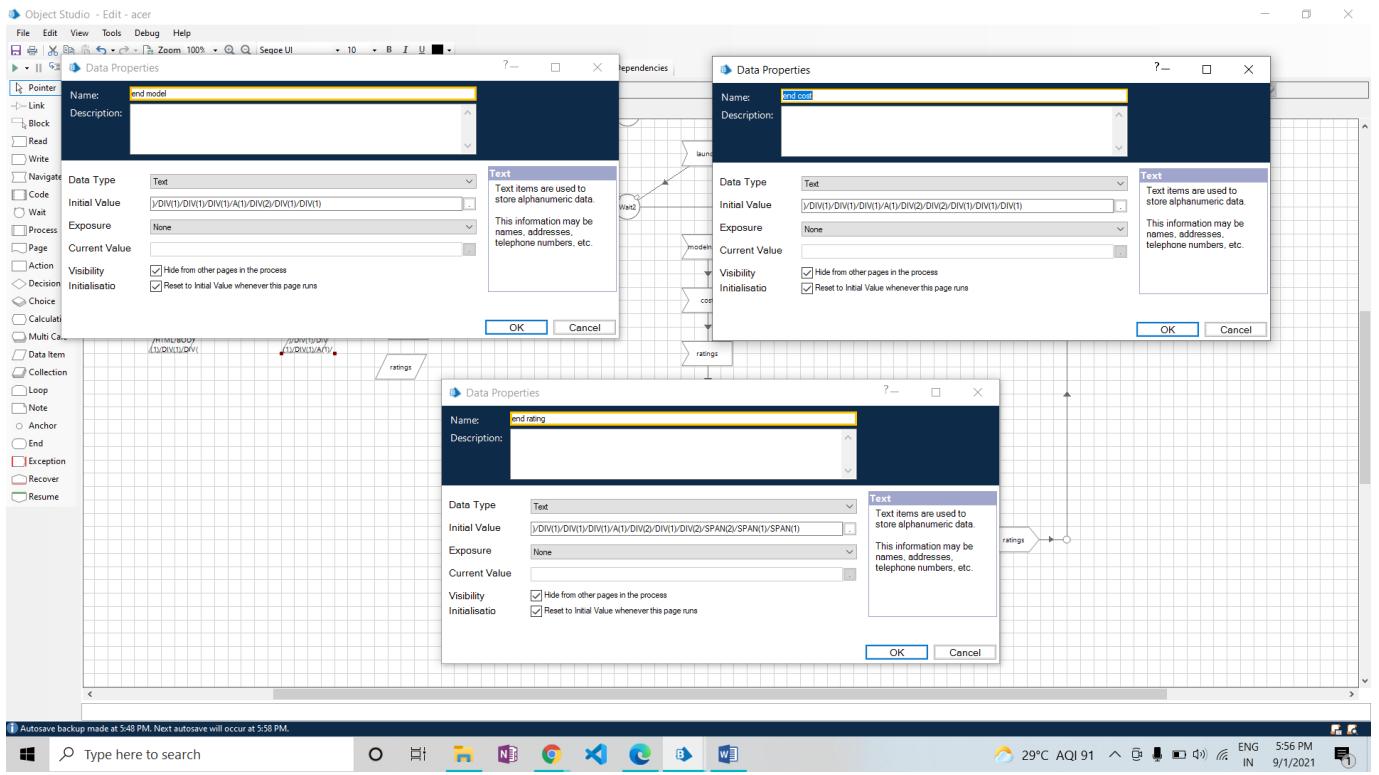












Screenshot of a web browser showing the Flipkart search results for "acer predator". The results page displays three Acer Predator Helios laptops with their prices, ratings, and brief descriptions.

Product	Price	Ratings
acer Predator Helios 300 Octa Core i7 10th Gen - (16 GB/1 TB HDD/256 GB SSD/Windows 10 Home/6 GB Graph... 4.5 ★ 585 Ratings & 104 Reviews	₹1,09,990	₹149,999 26% off
acer Predator Helios 300 Core i7 10th Gen - (16 GB/1 TB SSD/Windows 10 Home/6 GB Graphics/NVIDIA GeFor... 4.2 ★ 17 Ratings & 7 Reviews	₹1,19,990	₹166,000 27% off
acer Predator Helios 300 Core i5 7th Gen - (8 GB/1 TB HDD/128 GB SSD/Windows 10 Home/4 GB Graphics/NVI... 4.5 ★ 3,226 Ratings & 698 Reviews	₹1,29,999	₹149,999 25% off

Screenshot of Object Studio showing a process flow for extracting Acer laptop data from a web page. A collection properties dialog is open, listing the extracted fields and their current values.

Field	Initial Value	Current Value
modelname (Text)	cost (Text)	ratings (Text)
acer Predator Helios...	₹1,09,990	585 Ratings
acer Predator Helios...	₹1,19,990	17 Ratings
acer Predator Helios...	₹1,29,999	3,226 Ratings

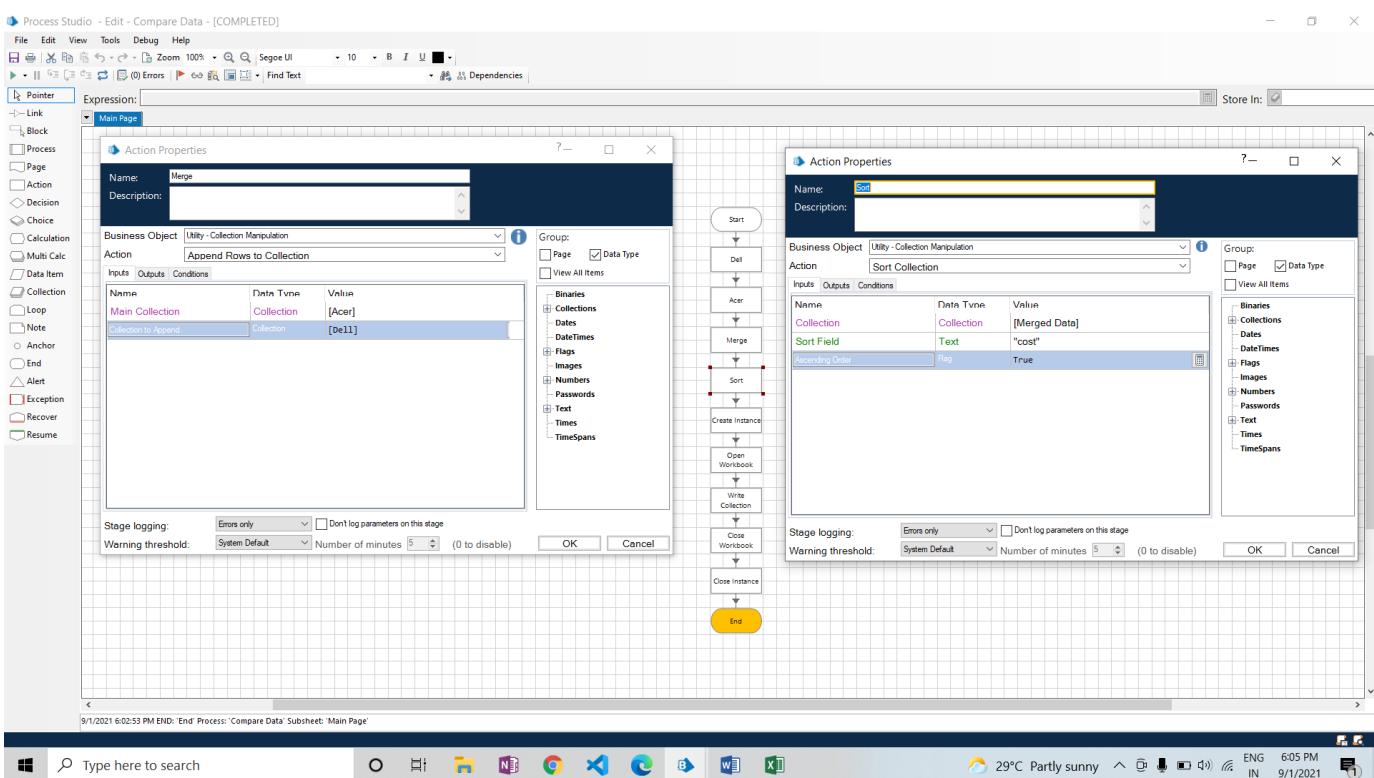
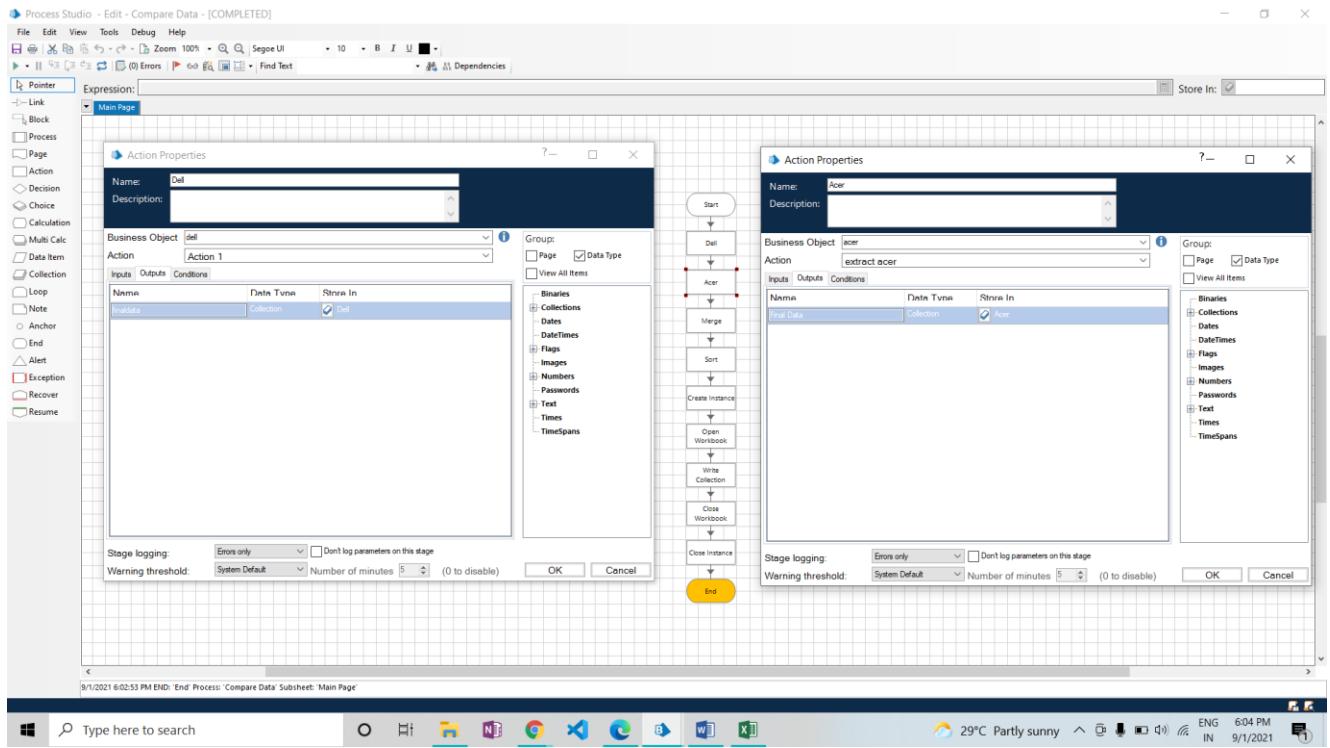
The process flow diagram shows a decision point based on the 'cost' field, leading to 'model', 'cost', and 'ratings' steps, which then converge back to a final output.

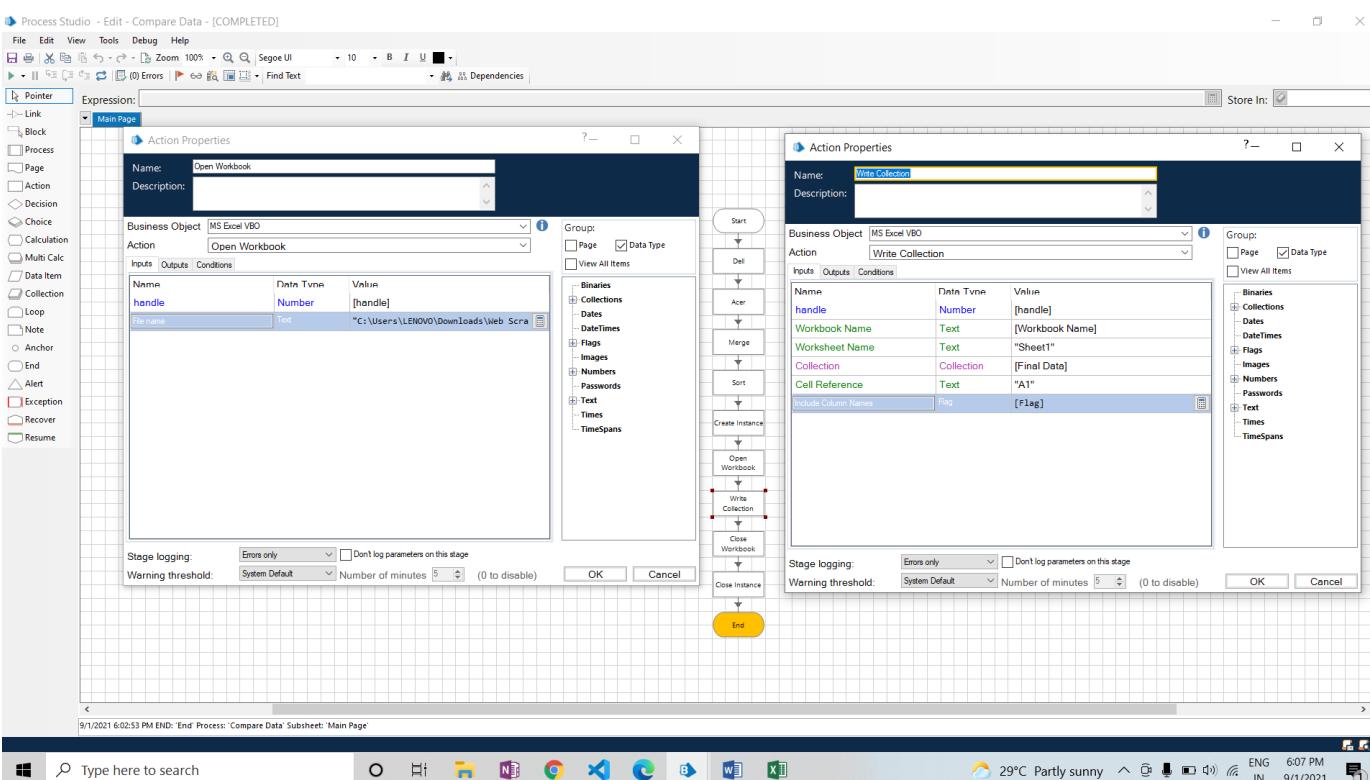
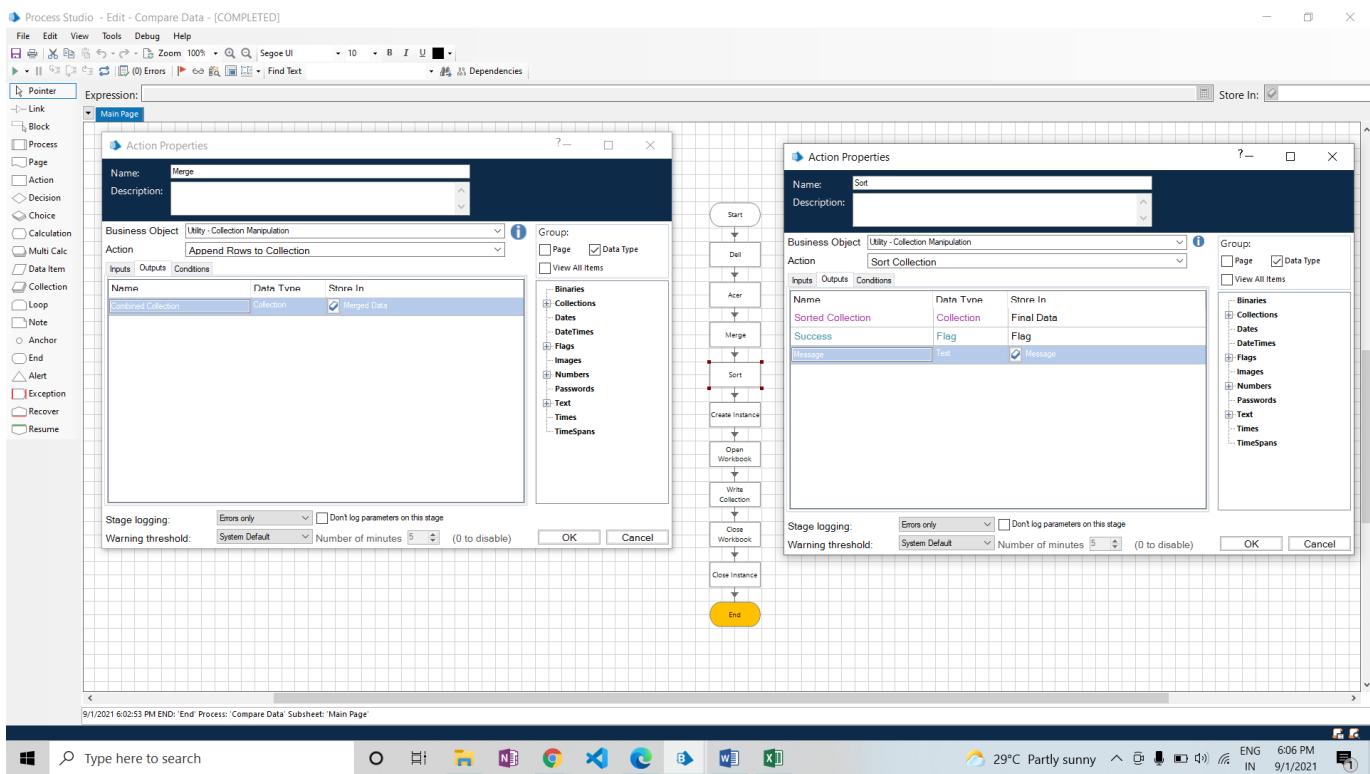
Screenshot of a web browser showing the Flipkart search results for "dell laptop". The search bar at the top contains the query "dell laptop". The results page shows three Dell laptop models: DELL Inspiron Core i5 11th Gen - (8 GB/512 GB SSD/Windows 10 Home/2 GB Graphics) Inspiron 5502 Thin and... (₹70,990), DELL Vostro Core i3 10th Gen - (8 GB/256 GB SSD/Windows 10 Home) Vostro 3401 Thin and Light Laptop (₹39,190), and DELL Inspiron Core i5 11th Gen - (8 GB/512 GB SSD/Windows 10) Inspiron 5410 2 in 1 Laptop (₹69,450). The left sidebar includes filters for Categories (Computers, Laptops), Price (₹75000+), Processor (Core i5, i3, i7, Ryzen 7, 5, 3, 5 Hexa, Dual Core), and Brand (DELL). The bottom status bar shows the URL as https://www.flipkart.com/dell-inspiron-core-i5-11th-gen-8-gb-512-gb-ssd-windows-10-home-2-graphics-5502-thin-light-laptop/p/itmfd763...

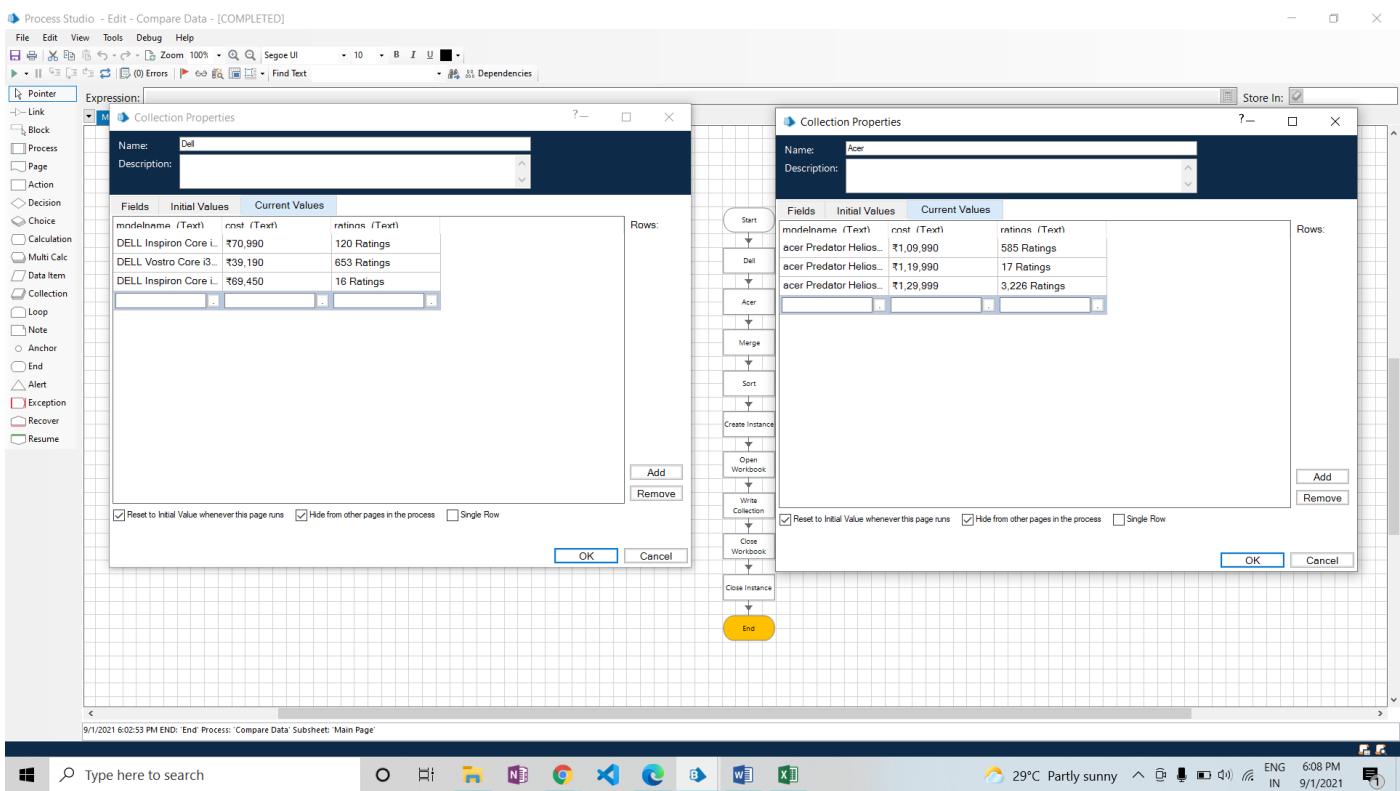
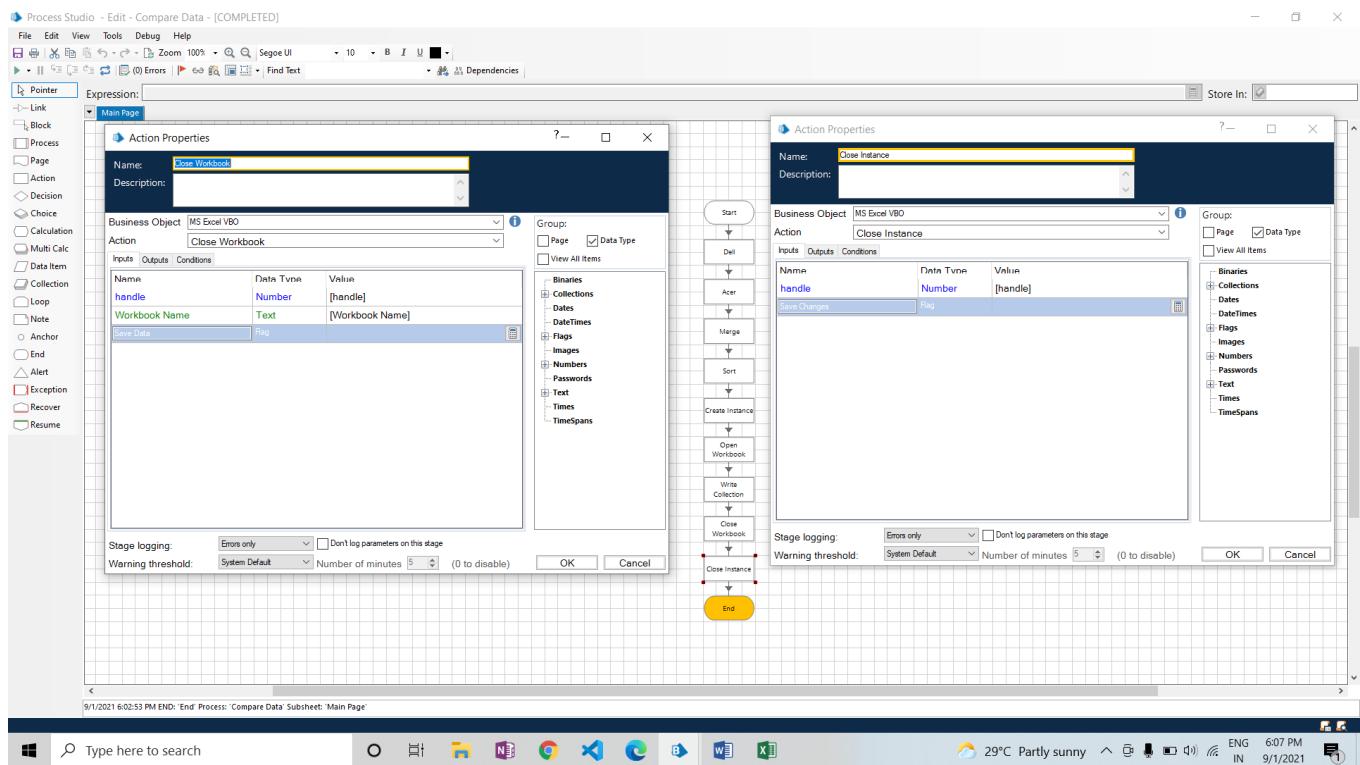
Screenshot of Object Studio software showing a process flow for extracting data from the Dell laptop search results. The main workspace displays a process diagram with various nodes like "Initialise", "Action 1", and "Decision1". A "Collection Properties" dialog box is open, showing a table with three rows of data:

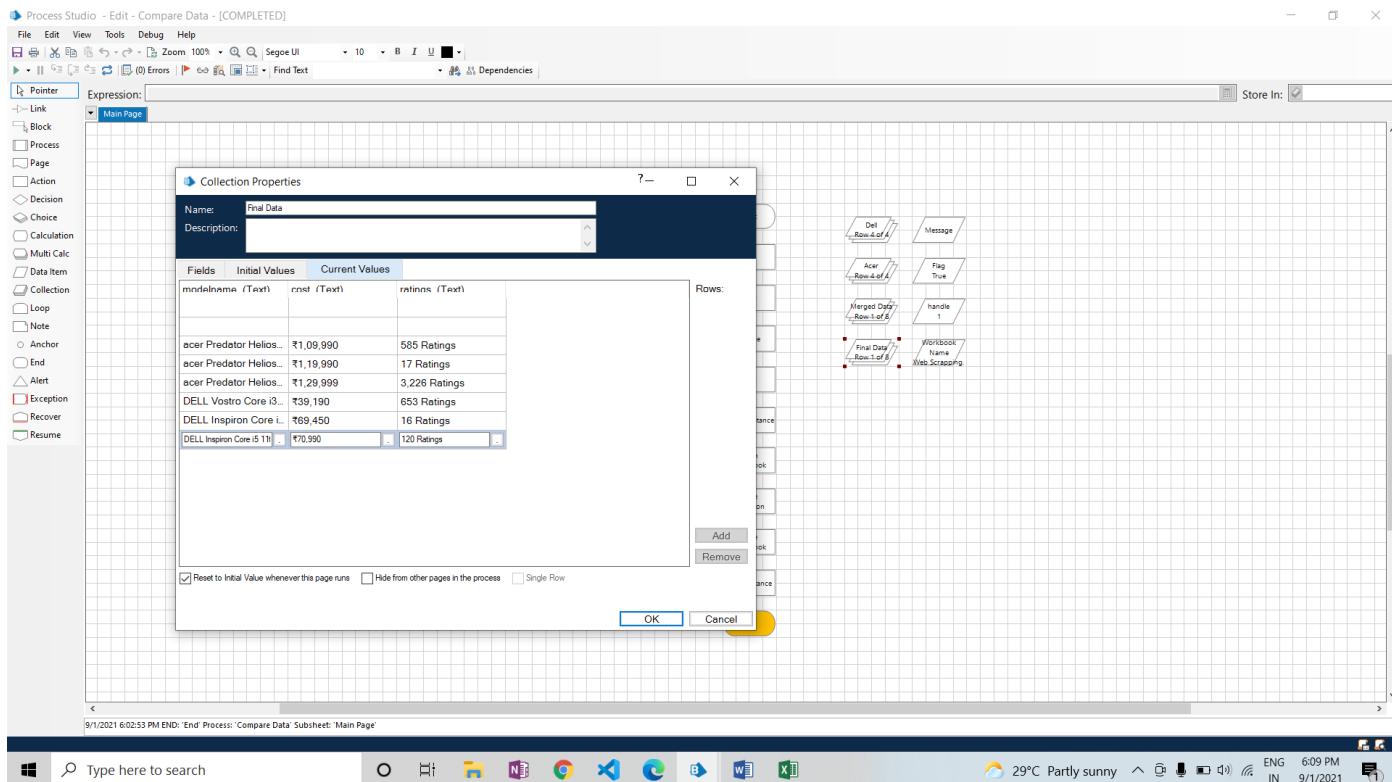
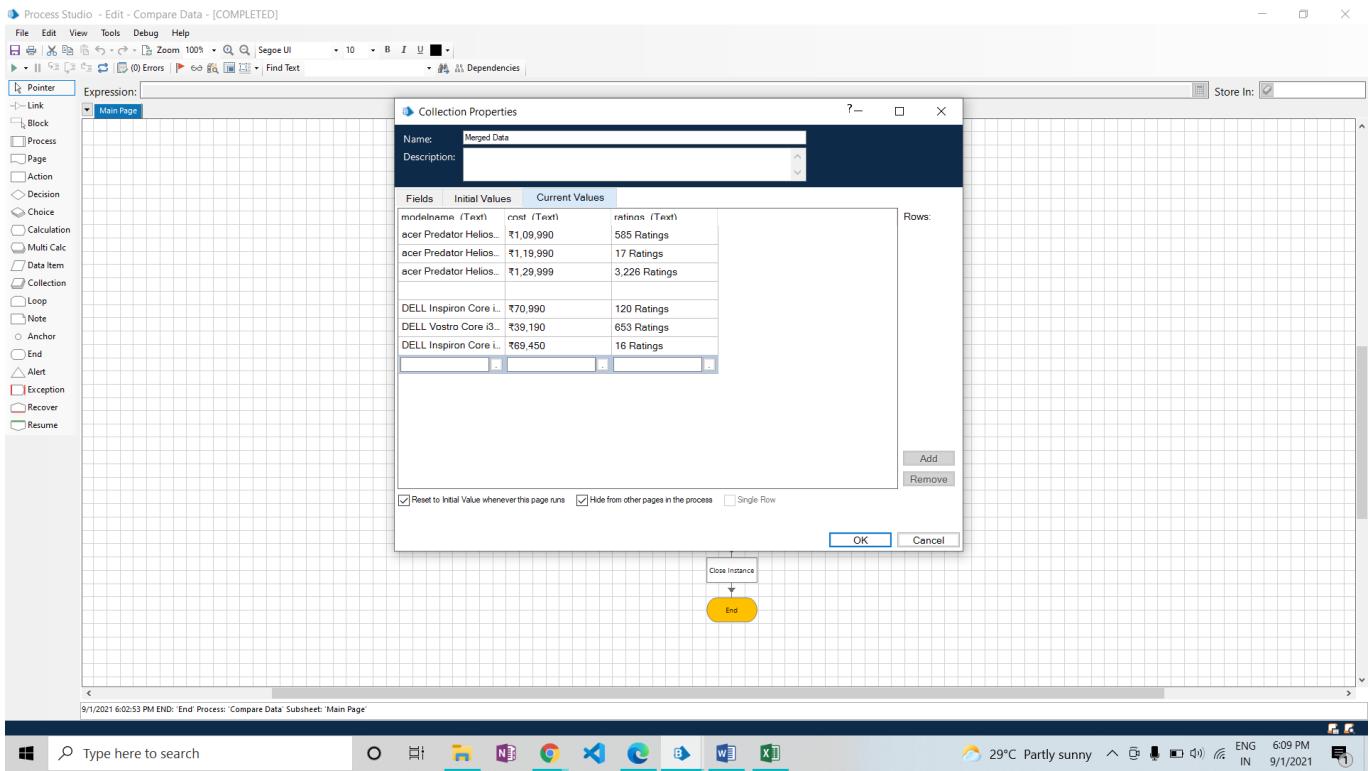
Fields	Initial Values	Current Values
modelname (Text)	cost (Text)	ratings (Text)
DELL Inspiron Core i...	₹70,990	120 Ratings
DELL Vostro Core i3...	₹39,190	653 Ratings
DELL Inspiron Core L...	₹69,450	16 Ratings

The bottom status bar shows the date and time as 9/1/2021 6:01:17 PM.









Web Scrapping - Excel (Product Activation Failed)

modelname	cost	ratings
4 acer Predator Helios 300 Octa Core i7 10th Gen - (16 GB/1 TB HDD/256 GB SSD/Windows 10 Home/6 GB Graph... 5 acer Predator Helios 300 Core i7 10th Gen - (16 GB/1 TB SSD/Windows 10 Home/6 GB Graphics/NVIDIA GeForce... 6 acer Predator Helios 300 Core i5 7th Gen - (8 GB/1 TB HDD/128 GB SSD/Windows 10 Home/4 GB Graphics/NVidia...	₹1,09,990 ₹1,19,990 ₹1,29,999	585 Ratings 17 Ratings 3,226 Ratings
7 DELL Vostro Core i3 10th Gen - (8 GB/256 GB SSD/Windows 10 Home) Vostro 3401 Thin and Light Laptop 8 DELL Inspiron Core i5 11th Gen - (8 GB/512 GB SSD/Windows 10) Inspiron 5410 2 in 1 Laptop 9 DELL Inspiron Core i5 11th Gen - (8 GB/512 GB SSD/Windows 10 Home/2 GB Graphics) Inspiron 5502 Thin an...	₹39,190 ₹69,450 ₹70,990	653 Ratings 16 Ratings 120 Ratings
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		

CHAPTER 6

CONCLUSION

In this project we have explored the Blue Prism concepts and learnt all concepts from the basics and implemented them in this project. For any project we do we must always first get an over view of what should be done for accomplishing the project, what are the requirements for the project as well.

Only after we were quite familiar with the concept through a little practice and syntax analysis, we started through our basic design, where we discovered additional concepts, which we needed to enhance our project. We did discover new functions we can use which can be called from the main function and executed.

There have been lots of steps and processes to follow, and many decision making mechanism are involved as well. But, we have tried our best to cover multiple scenarios and possibilities you might encounter. There are still challenges that could be out of the box, but this guidebook will hopefully help you create the basic framework for the process as well as the application. We would also like to thank our guide Ms. Sonali Dash and Ms. Sowmya BK to help us out teaching us all the concepts we needed to do the project through classes and sessions.

CHAPTER 7

FUTURE ENHANCEMENTS

This project is done using the basic concepts of the Robotic Process Automation. So, this project is implemented using the basic concepts we learnt in the possible time. We have used only few keys and functions for interacting. Therefore, we can further improve enhance the project by adding additional components into the project.

We can allot new functions such as where user can search their required item, where user can collect data from the website of their choice and sort the collected data according to their requirements. Also additionally we can allow users to collect, compare and store data of more than two items. Currently we have taken only two pre determined data from a single website who's data is collected and sorted according to the least cost , which further stored excel sheet for reference .

This project we have completed is used to show a basic concept of such a tedious and time consuming task can be automated very easily using RPA. The further enhancement which can made to our project will make the job of collecting and comparing data of an item from an e-commerce platform much more simpler, easier and faster. This technology can be put to use in various sectors and industries with slight modification according to their needs . This will be helpful to make lives much faster and better for everyone.

CHAPTER 8

REFERENCES

- 1.www.blueprism.com
- 2.<https://community.blueprism.com/>
- 3.https://www.youtube.com/results?search_query=data+extract+from+flipkart