Power Automate Web Scraper Documentation

Project: Car Dealership Vehicle Pricing

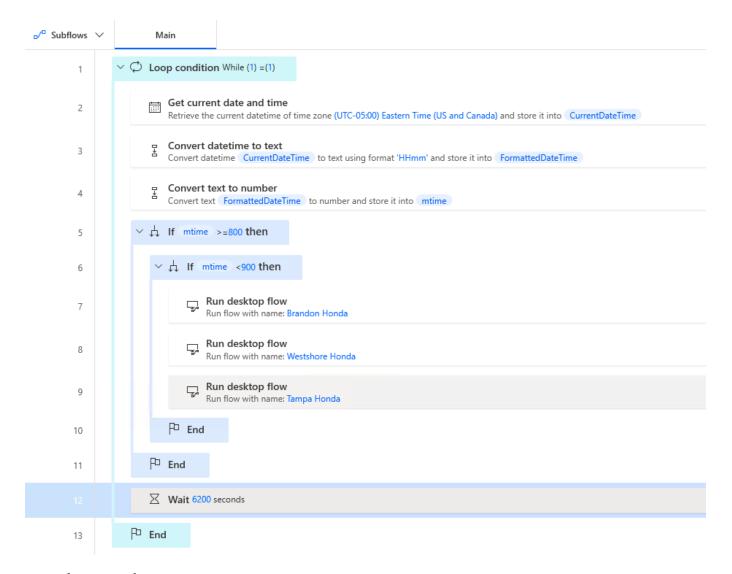
This document will cover the three different web scrapers that are used to extract data from three different Honda dealerships within the Tampa Bay area. The scrapers are similar in structure since they are built from a base template in order to quickly be deployed. Setting up a scraper for a new dealership can take as little as 30 minutes. The scrapers were intentionally changed to demonstrate different techniques for automating flows.

If you want to skip to the most interesting part of the flow that I have not seen anyone else discover, skip to section 8 in the Brandon Honda section.

Run Schedule

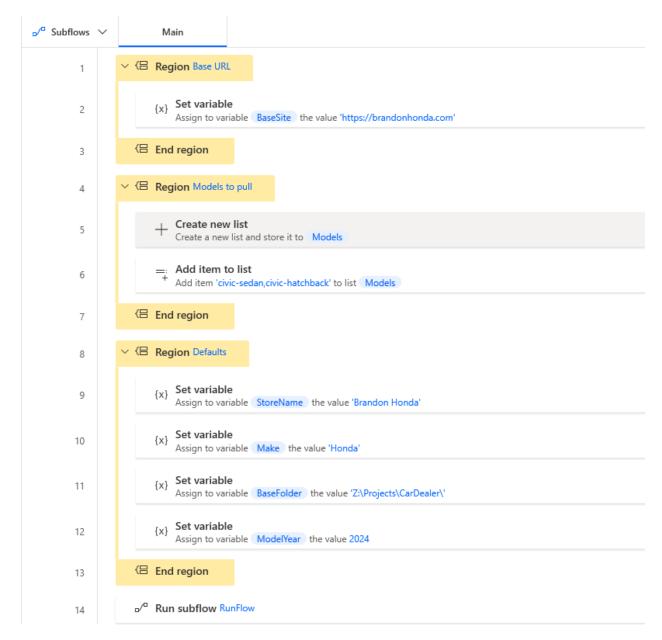
Currently I do not have access to Power Automate Cloud as I am trying to prevent spending money on this project, so the scrapers run on a schedule within a desktop flow. The desktop flow is running on a VM I created in Hyper-V on my server running on Windows Server 2019. It writes the files to the mapped network drive used for the entire project.

It only has a Main flow as it is a short flow. The loop condition is always true so it never stops. It gets the current date and time, and converts to military time. It then checks if it is between 800 and 900 (8:00 A.M.), then waits and checks the time again every 60 minutes.



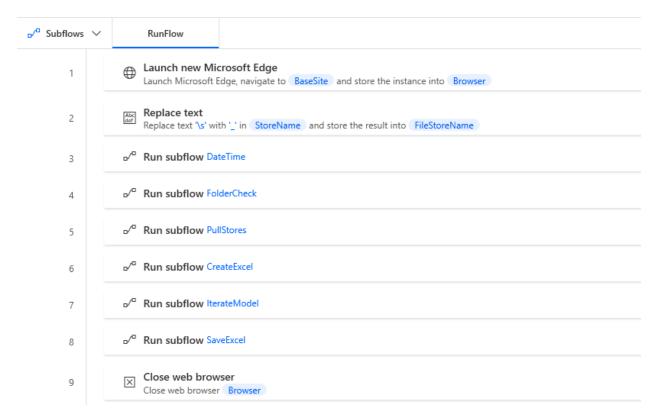
Brandon Honda

- 1. Main
 - a. The main flow is used to set variables, and add items to lists that will be used throughout the entire flow.



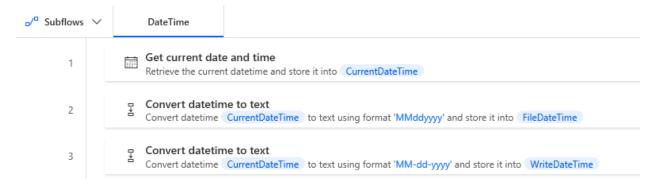
2. RunFlow

a. Next is a sub flow wrapper where I set all the sub flows to run in a specified order, and additionally place a few additional steps. In this case the web browser is launched, and the StoreName "Brandon Honda" will be set to "Brandon_Honda" as a different variable for saving the file. This will make the file easier to work with for setting it in a path.



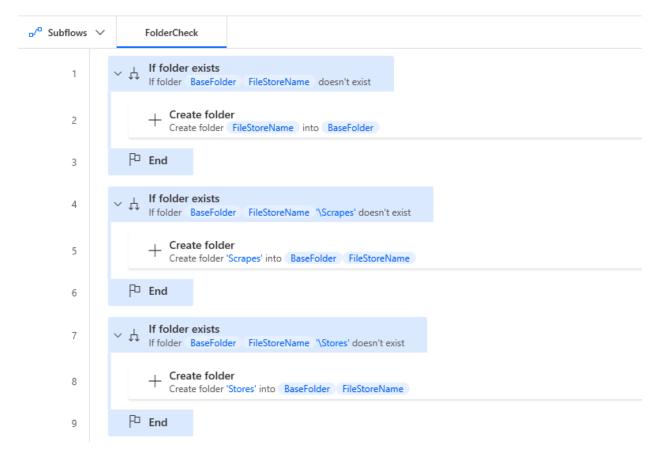
3. DateTime

a. This is just a short sub flow to get the date and setting it as a variable for the data table and the file name. The end file will use the store name and date when it is saved.



4. FolderCheck

a. Checks to see if directories exist for the store the scraper is being setup for, and creates them if not. These directory paths are where generated files will be saved to. This saves time when it comes to setting up file structures.



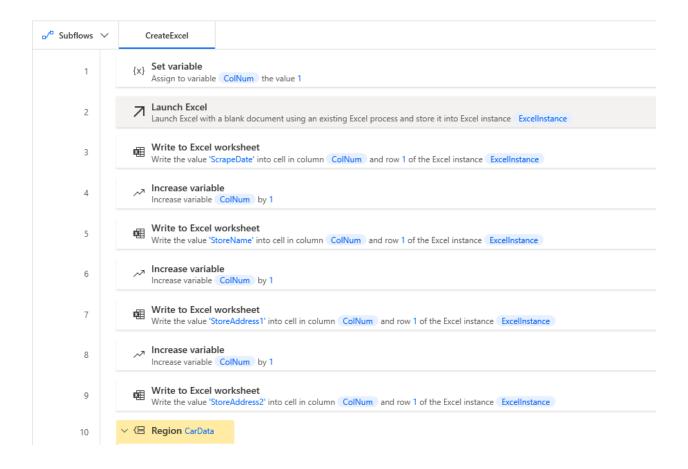
5. PullStores

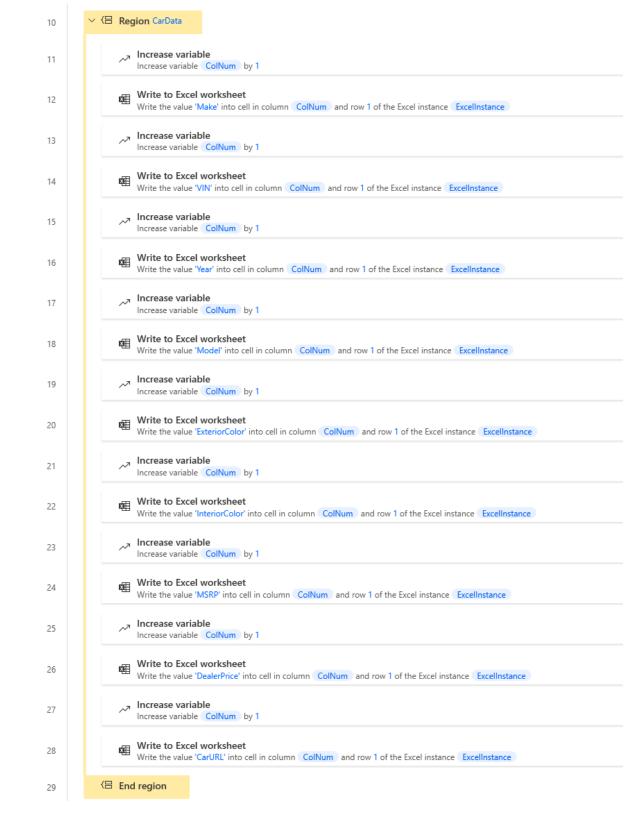
a. For this scraper, the store data is pulled from an excel file. The store URL is pulled from this file which will be iterated over, and the store details such as the name and address will be part of the output file.



6. CreateExcel

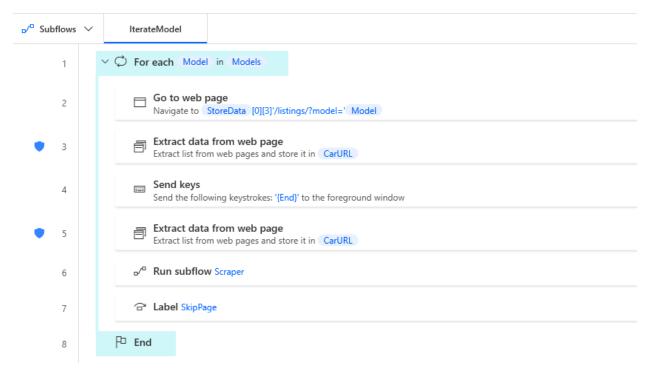
- a. Excel is the best method for writing data that is extracted using power automate. For SQL server you must iterate over each row of data. Even though the write speeds are fast, it is much quicker to write the entire data table to excel and set up an ETL or ELT package.
- b. This is broken into two parts for organization. The ungrouped part is for the data and store variables, while the grouped section will be for the extracted data. It's important to set this up to have the same number of columns that will be pulled as this will set your column headers. This usually comes after determining what data will be extracted.





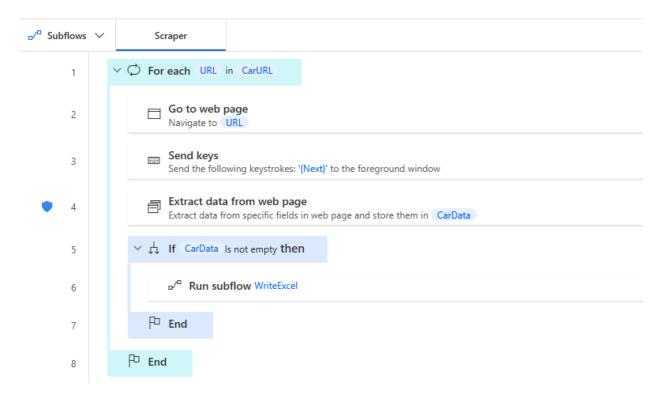
7. IterateModel

- a. The vehicle model variable is used in a for loop to iterate through the webpage listing. The store URL is used from the index indicated in the StoreData variable. This is combined with the rest of the URL needed to iterate through each model. For now it is only Honda Civics, but it could be expanded to pull other models.
- b. Full vehicle details cannot all be extracted from the main page listing for civics, so the URL link to each vehicle is pulled from the web page. Some of the URLs are not fully loaded unless the entire web page is scrolled to the bottom, so an end keystroke is simulated to go to the bottom of the page and load in all the URLs.

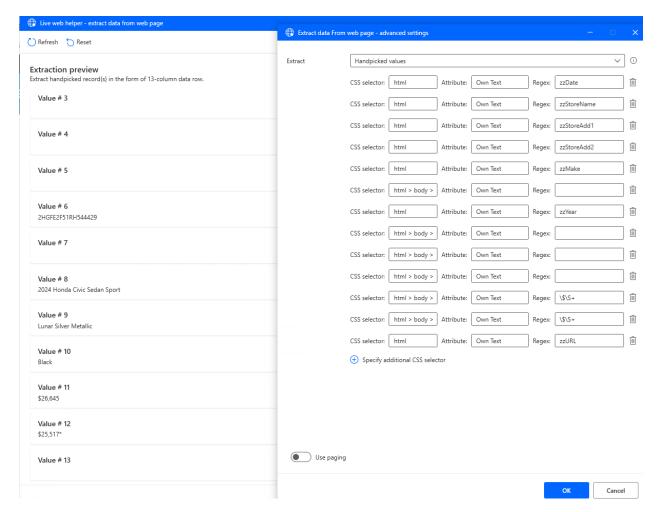


8. Scraper

- a. After the URLs for each vehicle are pulled, another for loop is used to go to each web page.
- b. The page down (Next) key is used to load in all vehicle details, and the data is extracted into CarData.

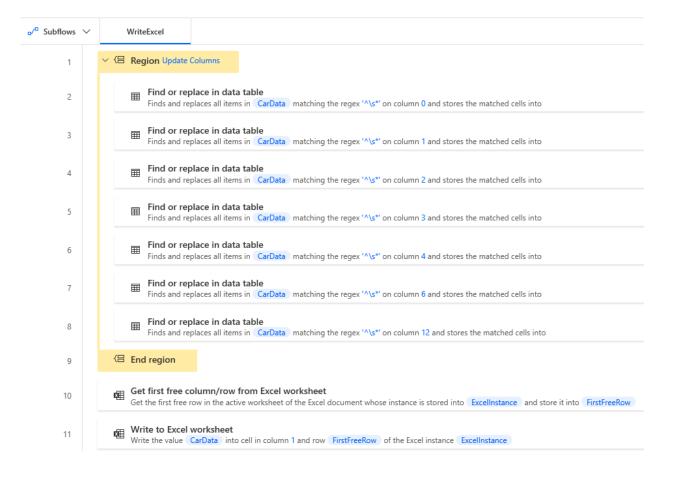


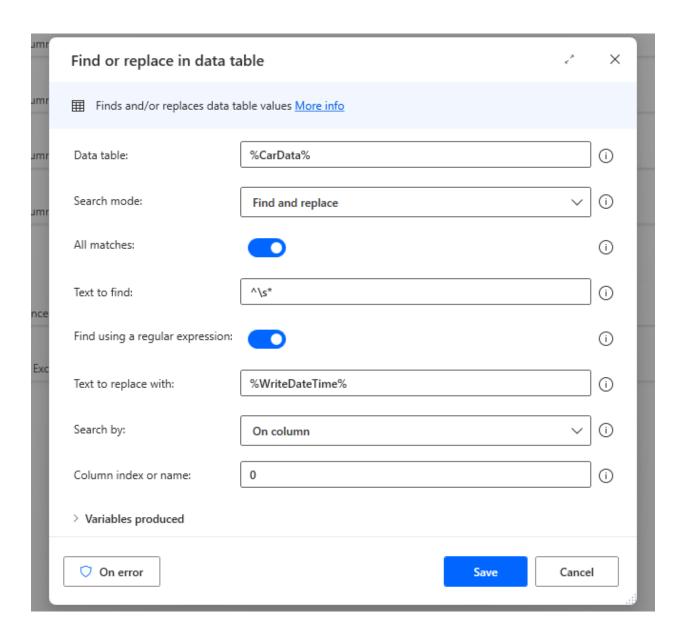
c. One thing I do that I have not seen anyone else discover after reading through many different forms and threads is how I set up the data table to be filled with variables. I specify additional CSS selectors until I have the same number of columns as created in the excel file. Then I used the base part of the CSS selector paths for all extracted elements, set the attribute to Own Text, and then use a regular expression that wont pull any data. This always results in an empty column which can be filled with values from a variable or specified otherwise.



9. WriteExcel

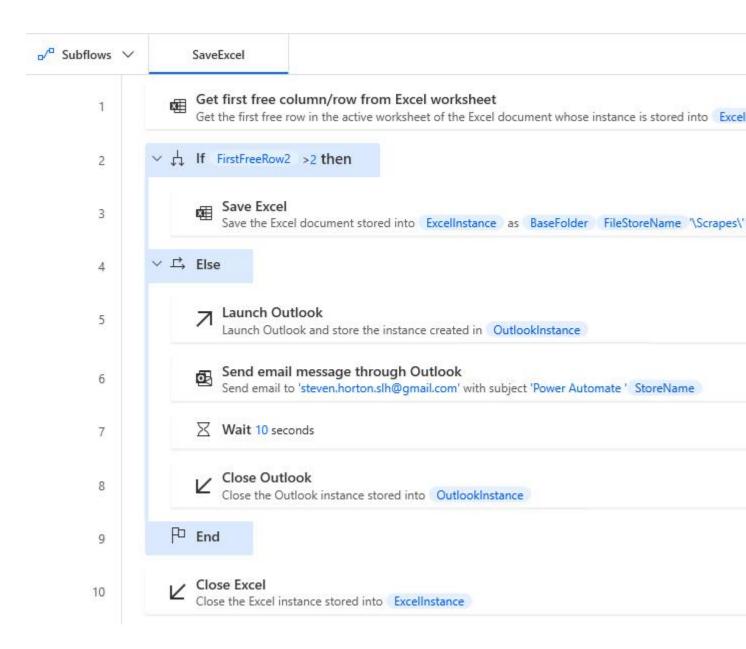
- a. This is where the data is written to the Excel file. First, find and replace is used on the data table to fill the empty column with values. The regular expressions "^\s " is used to find empty fields in the column index specified, which is then replaced with the specified variable (%WriteDateTime%).
- b. Next the first free row is determined, and the data is written on that row

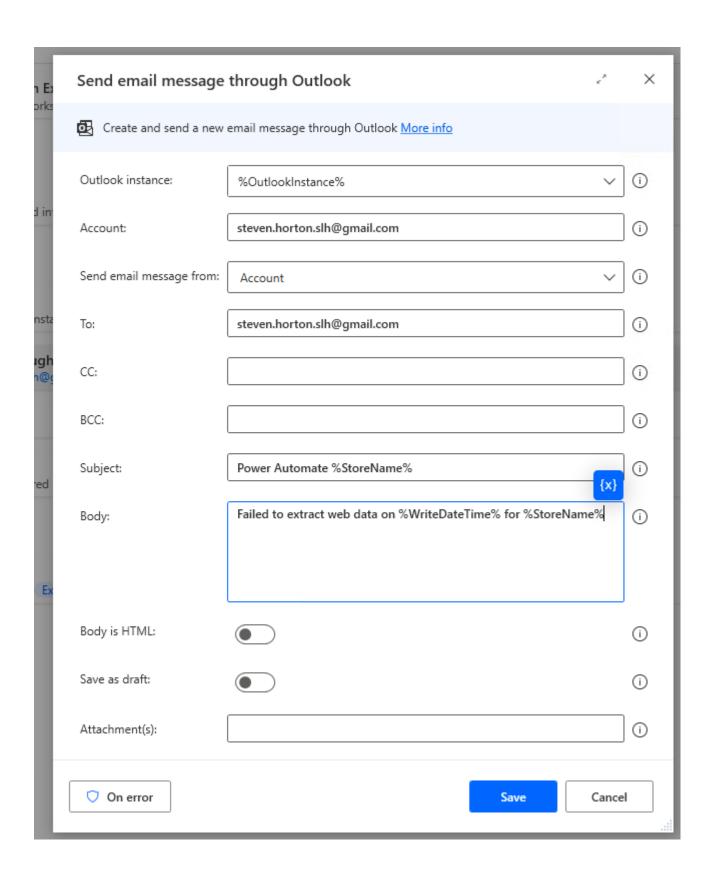




10. SaveExcel

- a. After all the vehicles and models have been iterated over, the Excel file is saved to the directory location with the variables indicating the store and date.
- b. If for any reason no data is extracted and the Excel file is empty, an email will be sent as a notification of the failed data extraction for the data and store.
- c. The Excel file shown is what the final output looks like.





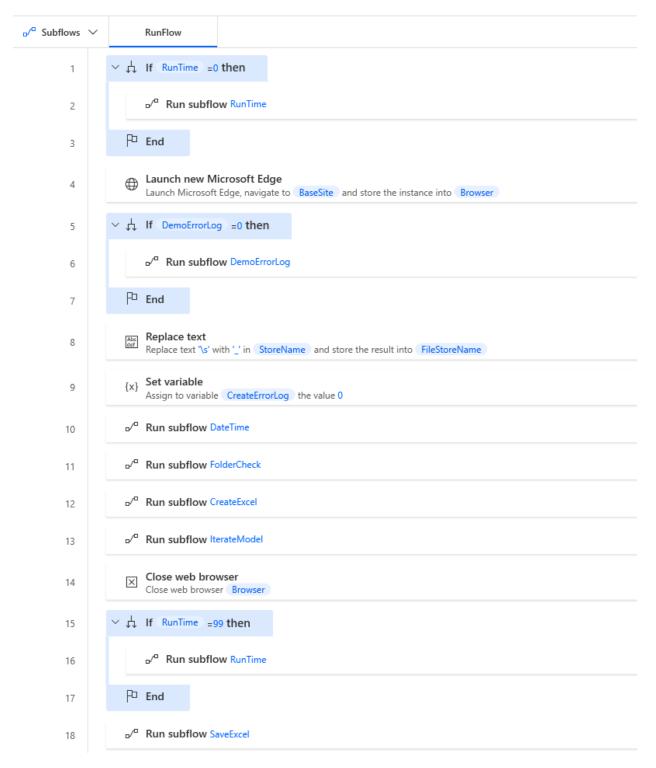
	100		D	110720		1.21	H			· · ·		M
Α	StoreName	C StoreAddress1		E	F	G Year Model	н	ExteriorColor	InteriorColo		DealerPrid	
					19XFL2H59RE015467		la Civic Hatchback LX	Meteorite Gray Metallic			\$25,075*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-lx-19xfl2h59re015467-e015467/
					19XFL2H56RE018956		la Civic Hatchback LX	Platinum White Pearl	Black		\$25,505*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-lx-19xf12h56re018956-e018956/
					19XFL2H57RE024605		la Civic Hatchback LX	Platinum White Pearl	Black		\$25,505*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-lx-19xf12h57re024605-e024605/
					2HGFE2F51RH572604		la Civic Sedan Sport	Rallye Red	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f51rh572604-h572604/
					2HGFE2F59RH572592			Rallye Red	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f59rh572592-h572592/
					2HGFE2F55RH572089			Meteorite Gray Metallic				https://brandonnonda.com/listings/new-2024-honda-civic-sedan-sport-2ngre2r59rh572089-h573089/ https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f55rh573089-h573089/
											\$25,517*	
					2HGFE2F55RH553165			Rallye Red	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f55rh553165-h553165/
					2HGFE2F50RH571671			Crystal Black Pearl	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f50rh571671-h571671/
					2HGFE2F52RH573552			Crystal Black Pearl	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f52rh573552-h573552/
					2HGFE2F5XRH573640			Aegean Blue Metallic	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f5xrh573640-h573640/
					2HGFE2F51RH544429			Lunar Silver Metallic	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f51rh544429-h544429/
					2HGFE2F56RH573036			Meteorite Gray Metallic			\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f56rh573036-h573036/
					2HGFE2F57RH562367			Lunar Silver Metallic	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f57rh562367/
					2HGFE2F54RH572029			Meteorite Gray Metallic			\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f54rh572029-h572029/
					2HGFE2F50RH564820			Platinum White Pearl	Black		\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f50rh564820-h564820/
					2HGFE2F56RH565194		la Civic Sedan Sport	Platinum White Pearl	Black		\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f56rh565194-h565194/
					2HGFE2F59RH564816			Platinum White Pearl	Black		\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f59rh564816-h564816/
					2HGFE2F5XRH560497		la Civic Sedan Sport	Platinum White Pearl	Black		\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f5xrh560497-h560497/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL2H83RE025843	2024 2024 Hono	la Civic Hatchback Sport	Lunar Silver Metallic	Black	\$27,445	\$26,280*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h83re025843-e025843/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL2H83RE025082	2024 2024 Hono	la Civic Hatchback Sport	Rallye Red	Black	\$27,445	\$26,280*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h83re025082-e025082/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL2H86RE025836	2024 2024 Hono	la Civic Hatchback Sport	Lunar Silver Metallic	Black	\$27,445	\$26,280*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h86re025836-e025836/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL2H89RE020047	2024 2024 Hono	la Civic Hatchback Sport	Platinum White Pearl	Black	\$27,900	\$26,711*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h89re020047-e020047/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL2H8XRE018727	2024 2024 Hono	la Civic Hatchback Sport	Platinum White Pearl	Black	\$27,900	\$26,711*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h8xre018727-e018727/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL1H77RE008724	2024 2024 Hono	la Civic Hatchback EX-L	Meteorite Gray Metallic	Black	\$29,745	\$28,446*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h77re008724-e008724/
4/24/2024	Brandon Honda	9209 E Adamo De	Tampa, FL 3361	9 Honda	19XFL1H71RE004765	2024 2024 Hono	la Civic Hatchback EX-L	Meteorite Gray Metallic	Gray	\$29,745	\$28,473*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h71re004765-re004765/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL1H78RE005931	2024 2024 Hono	la Civic Hatchback EX-L	Meteorite Gray Metallic	Gray	\$29,745	\$28,473*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h78re005931-e005931/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	19XFL1H77RE019884	2024 2024 Hono	la Civic Hatchback EX-L	Smokey Mauve Pearl	Black	\$30,200	\$28,903*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h77re019884-e019884/
4/24/2024	Brandon Honda	9209 E Adamo De	Tampa, FL 3361	9 Honda	19XFL1H78RE000843	2024 2024 Hono	la Civic Hatchback EX-L	Platinum White Pearl	Black	\$30,200	\$28,903*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h78re000843-e000843/
4/24/2024	Brandon Honda	9209 E Adamo De	Tampa, FL 3361	9 Honda	19XFL1H72RE001213	2024 2024 Hono	la Civic Hatchback EX-L	Platinum White Pearl	Black	\$30,200	\$28,903*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h72re001213-e001213/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	2HGFE1F92RH303556	2024 2024 Hono	la Civic Sedan Touring	Meteorite Gray Metallic	Gray	\$31,645	\$29,780*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-touring-2hgfe1f92rh303556-h303556/
4/24/2024	Brandon Honda	9209 E Adamo De	Tampa, FL 3361	9 Honda	2HGFE1F92RH328604	2024 2024 Hono	la Civic Sedan Touring	Crystal Black Pearl	Black	\$31,645	\$29,809*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-touring-2hgfe1f92rh328604-h328604/
4/24/2024	Brandon Honda	9209 E Adamo Dr	Tampa, FL 3361	9 Honda	2HGFE1F96RH315533	2024 2024 Hono	la Civic Sedan Touring	Platinum White Pearl	Gray	\$32,100	\$30,239*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-touring-2hgfe1f96rh315533-h315533/
					2HGFE1F95RH310615		la Civic Sedan Touring	Platinum White Pearl	Black		\$30,239*	https://brandonhonda.com/fistings/new-2024-honda-civic-sedan-touring-2hgfe1f95rh310615-h310615/
4/24/2024	Brandon Honda	9209 E Adamo De	Tampa, FL 3361	9 Honda	19XFL1H8XRE017877	2024 2024 Hono	la Civic Hatchback Sport Touri	g Platinum White Pearl	Black	\$33,000	\$31,083*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-touring-19xfl1h8xre017877-e017877/
					19XFL1G89RE006953		la Civic Hatchback Sport Touri		Black	\$33,000	\$31,083*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-touring-19xfl1g89re006953-e006953/
			And the Control of th					Was entered to the first of the same				

Westshore Honda

I will keep it short and show the different features between this flow and the previous flow. This adds run time tracking and has an error log.

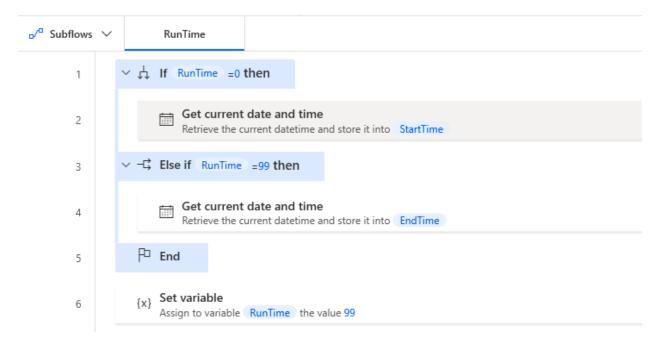
1. RunFlow

a. The RunTime variable is set to 0 in the Main flow. This triggers the sub flow through an if statement. RunTime is set to 99 in the RunTime sub flow, so it is triggered again at the very end of the entire flow.



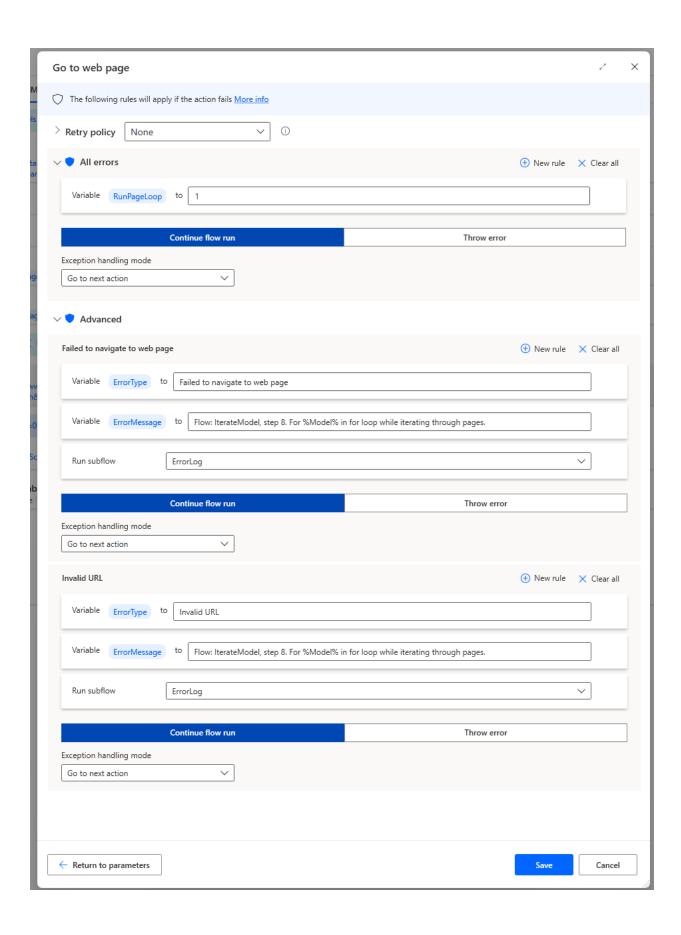
2. RunTime

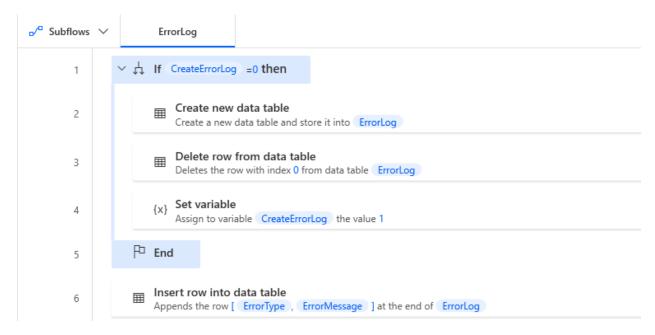
a. A simple sub flow to capture the start time at the beginning when the RunTime variable is set to 0. It then changes to 99 so once the flow is triggered again at the end it sets the time to EndTime instead of StartTime.



3. ErrorLog

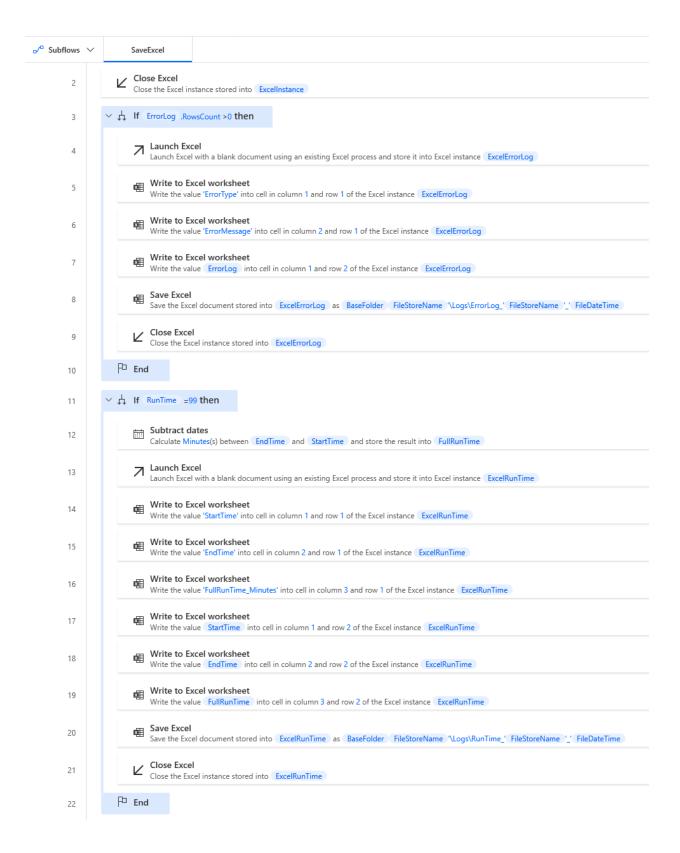
- a. An error log is created to track issues that happen within the flow. This allows you to continue running the rest of the flow if it is not a critical error. The error log is later written to an Excel file with all errors incurred in the flow.
- b. The errors are produced by setting error rules within a step in certain actions. The variables are set to be written in the ErrorLog sub flow, which is run whenever an error occurs.





4. SaveExcel

a. The error log and run time files are created and written in the SaveExcel sub flow, and written to their perspective locations within the directory structure.



As of right now the Tamp Honda scraper has no additional features or elements, but it will pull the stores from SQL, and setup directory locations and other variables from a config file in the future. The config file is something you can edit and have the changes apply to all flows that pull their config variables from the file, rather than hard coding each configuration and variables into the flow.