Power Automate Web Scraper Documentation

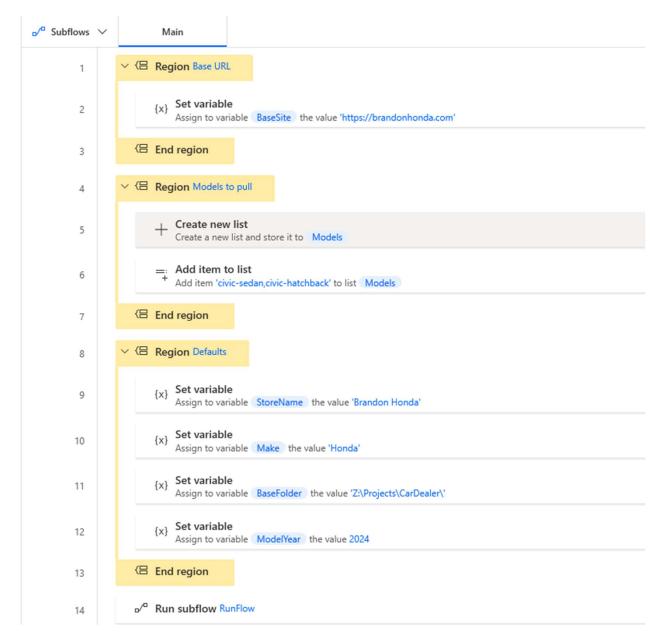
Project: Car Dealership Vehicle Pricing

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.

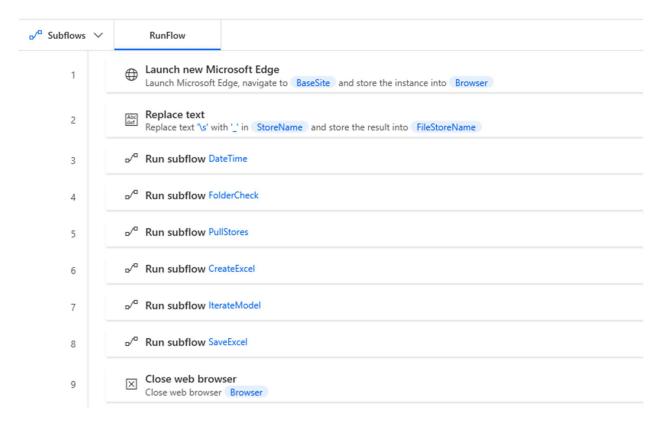
Brandon Honda

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



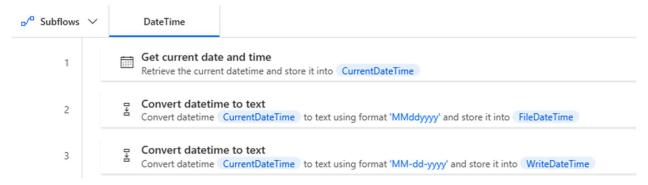
2. RunFlow

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.



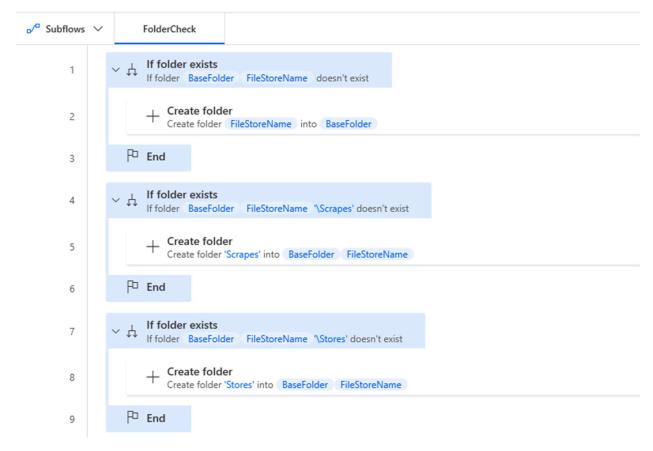
3. DateTime

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.



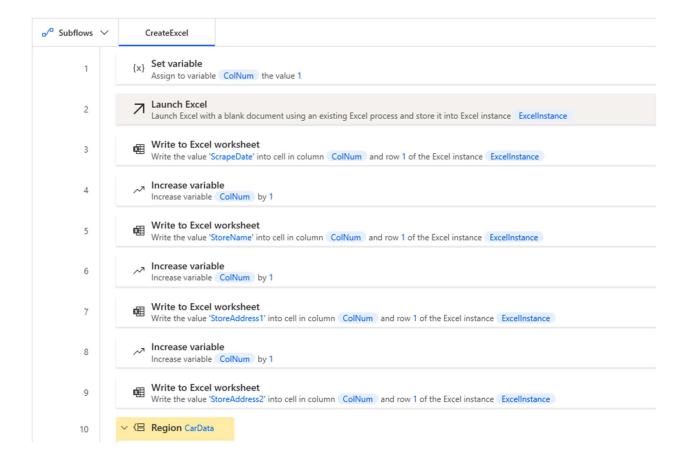
5. PullStores

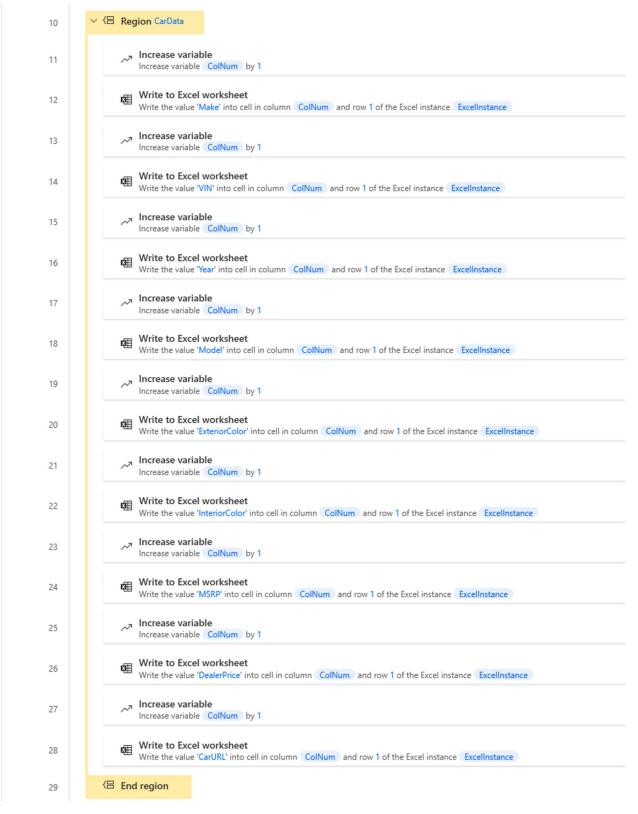
a. For this scraper, the store data is pulled from an excel file. The store URL is pulled from this file which will be iterated over, and the store details such as the name and address will be part of the output file.



6. CreateExcel

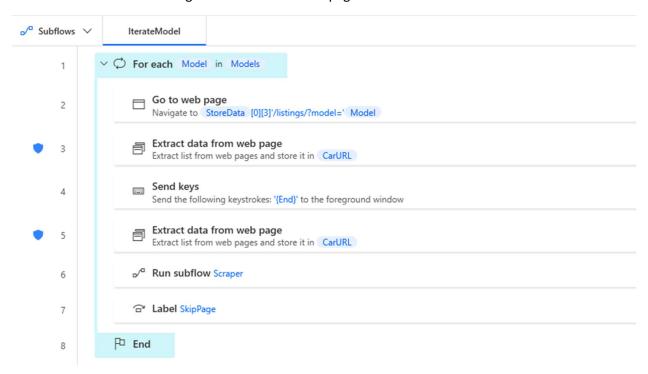
- a. Excel is the best method for writing data that is extracted using power automate. For SQL server you must iterate over each row of data. Even though the write speeds are fast, it is much quicker to write the entire data table to excel and set up an ETL or ELT package.
- b. This is broken into two 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.





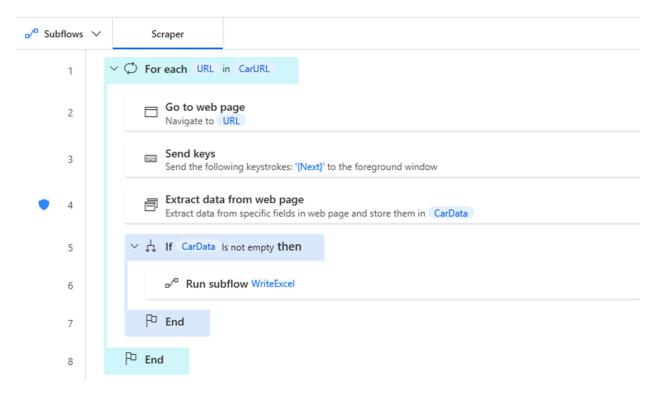
7. IterateModel

- 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.

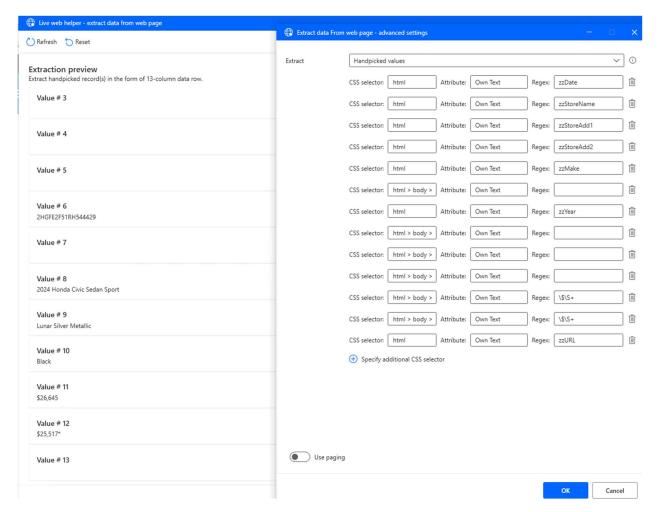


8. Scraper

- 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.

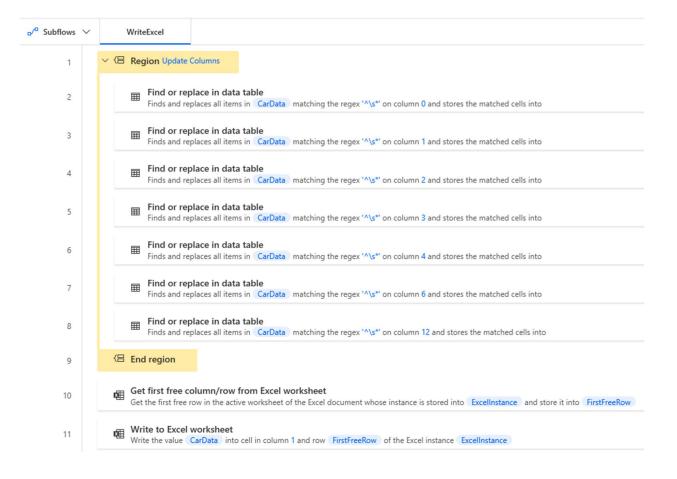


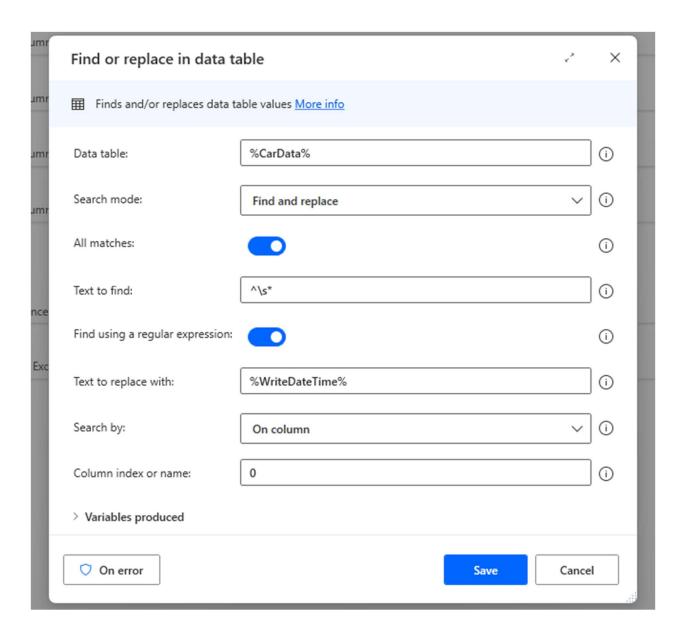
c. One thing I do that I have not seen anyone else discover after reading through many different forms and threads is how I set up the data table to be filled with variables. I specify additional CSS selectors until I have the same number of columns as created in the excel file. Then I used the base 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

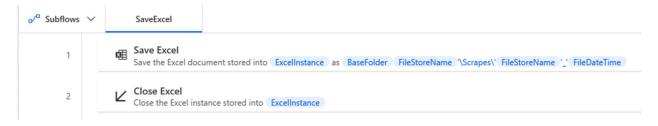
- a. This is where the data is written to the Excel file. First, find and replace is used on the data table to fill the empty column with values. The regular expressions "^\s " is used to find empty fields in the column index specified, which is then replaced with the specified variable (%WriteDateTime%).
- 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. The Excel file shown is what the final output looks like.



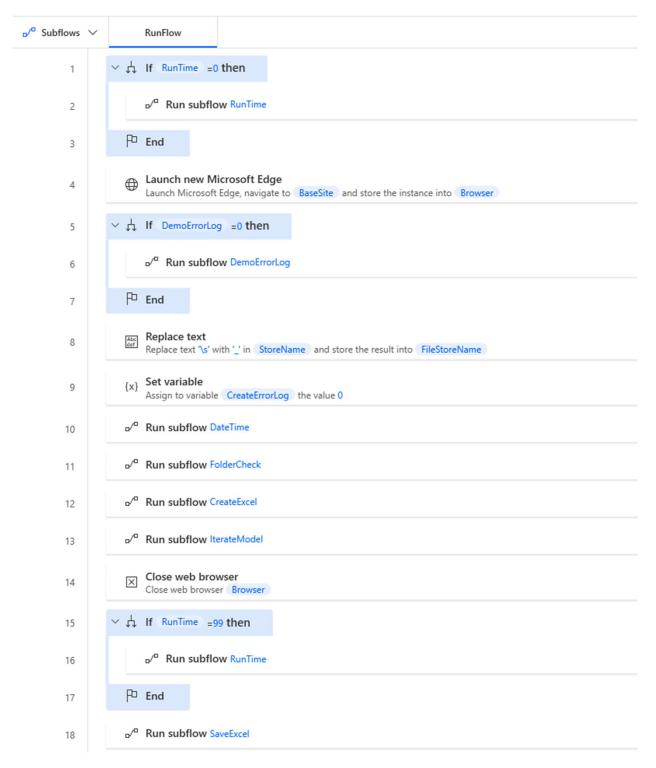
			12		1 2		2	-	H			K		
4	A	В	C	D	E		F	G		ExteriorColor	InteriorColor	- 11	DealerPrice	M
	crapeDate S			StoreAddress2 Tampa, FL 33619				Year Mi	odei 24 Honda Civic Hatchback LX	Meteorite Gray Metallic			\$25.075*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-lx-19xfl2h59re015467-e015467/
				Tampa, FL 33619					24 Honda Civic Hatchback LX	Platinum White Pearl	Black	\$26,500		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-lx-19xfl2h56re018956-e018956/
٠				Tampa, FL 33619					24 Honda Civic Hatchback LX	Platinum White Pearl	Black	\$26,500		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-lx-19xfl2h57re024605-e024605/
									24 Honda Civic Hatchback LX 24 Honda Civic Sedan Sport	Rallye Red	Black		\$25,505	https://brandonhonda.com/listings/new-2024-honda-civic-natchback-ix-19x12ris/re024003-e024003/ https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f51rh572604-h572604/
									24 Honda Civic Sedan Sport	Rallye Red	Black	\$26,645		https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2/ig/e2/51/h572592/
									24 Honda Civic Sedan Sport 24 Honda Civic Sedan Sport			\$26,645		https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f55rh573089-h573089/
									24 Honda Civic Sedan Sport 24 Honda Civic Sedan Sport	Meteorite Gray Metallic Rallye Red	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-nonda-civic-sedan-sport-2hgfe2f55rh553165-h553165/
									24 Honda Civic Sedan Sport	Crystal Black Pearl	Black		\$25,517*	
											Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f50rh571671-h571671/ https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f52rh573552/
									24 Honda Civic Sedan Sport	Crystal Black Pearl				
									24 Honda Civic Sedan Sport	Aegean Blue Metallic	Black	\$26,645		https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f5xrh573640-h573640/
									24 Honda Civic Sedan Sport	Lunar Silver Metallic			\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f51rh544429-h544429/
									24 Honda Civic Sedan Sport 24 Honda Civic Sedan Sport	Meteorite Gray Metallic Lunar Silver Metallic	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f56rh573036-h573036/ https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f57rh562367-h562367/
									24 Honda Civic Sedan Sport	Meteorite Gray Metallic	Black		\$25,517*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f54rh572029-h572029/
									24 Honda Civic Sedan Sport	Platinum White Pearl			\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f50rh564820-h564820/
									24 Honda Civic Sedan Sport	Platinum White Pearl	Black		\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f56rh565194-h565194/
									24 Honda Civic Sedan Sport	Platinum White Pearl	Black		\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f59rh564816-h564816/
									24 Honda Civic Sedan Sport	Platinum White Pearl	Black		\$25,948*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-sport-2hgfe2f5xrh560497-h560497/
				Tampa, FL 33619					24 Honda Civic Hatchback Sport	Lunar Silver Metallic	Black		\$26,280*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h83re025843-e025843/
				Tampa, FL 33619					24 Honda Civic Hatchback Sport	Rallye Red	Black	\$27,445		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xf12h83re025082-e025082/
				Tampa, FL 33619					24 Honda Civic Hatchback Sport	Lunar Silver Metallic	Black	\$27,445		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xf12h86re025836-e025836/
				Tampa, FL 33619					24 Honda Civic Hatchback Sport	Platinum White Pearl	Black	\$27,900		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h89re020047-e020047/
				Tampa, FL 33619					24 Honda Civic Hatchback Sport	Platinum White Pearl	Black		\$26,711*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-19xfl2h8xre018727-e018727/
				Tampa, FL 33619					24 Honda Civic Hatchback EX-L	Meteorite Gray Metallic		\$29,745		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h77re008724-e008724/
				Tampa, FL 33619					24 Honda Civic Hatchback EX-L	Meteorite Gray Metallic		\$29,745		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h71re004765-re004765/
				Tampa, FL 33619					24 Honda Civic Hatchback EX-L	Meteorite Gray Metallic		\$29,745		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h78re005931-e005931/
				Tampa, FL 33619					24 Honda Civic Hatchback EX-L	Smokey Mauve Pearl	Black	\$30,200		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h77re019884-e019884/
				Tampa, FL 33619					24 Honda Civic Hatchback EX-L	Platinum White Pearl	Black	\$30,200		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h78re000843-e000843/
				Tampa, FL 33619					24 Honda Civic Hatchback EX-L	Platinum White Pearl	Black	\$30,200		https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-ex-l-19xfl1h72re001213-e001213/
								2024 20	24 Honda Civic Sedan Touring	Meteorite Gray Metallic	Gray	\$31,645	\$29,780*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-touring-2hgfe1f92rh303556-h303556/
				Tampa, FL 33619					24 Honda Civic Sedan Touring	Crystal Black Pearl	Black	\$31,645	\$29,809*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-touring-2hgfe1f92rh328604-h328604/
									24 Honda Civic Sedan Touring	Platinum White Pearl	Gray	\$32,100		https://brandonhonda.com/listings/new-2024-honda-civic-sedan-touring-2hgfe1f96rh315533-h315533/
	4/24/2024 B	trandon Honda	9209 E Adamo Dr	Tampa, FL 33619	Honda	2HGFE1F9	95RH310615	2024 20	24 Honda Civic Sedan Touring	Platinum White Pearl	Black	\$32,100	\$30,239*	https://brandonhonda.com/listings/new-2024-honda-civic-sedan-touring-2hgfe1f95rh310615-h310615/
5	4/24/2024 B	Irandon Honda	9209 E Adamo Dr	Tampa, FL 33619	Honda	19XFL1H8	BXRE017877	2024 20	24 Honda Civic Hatchback Sport Touring	Platinum White Pearl	Black	\$33,000	\$31,083*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-touring-19xfl1h8xre017877-e01787
	4/24/2024 B	trandon Honda	9209 E Adamo Dr	Tampa, FL 33619	Honda	19XFL1G8	B9RE006953	2024 20	24 Honda Civic Hatchback Sport Touring	Sonic Gray Pearl	Black	\$33,000	\$31,083*	https://brandonhonda.com/listings/new-2024-honda-civic-hatchback-sport-touring-19xfl1g89re006953-e00695

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.

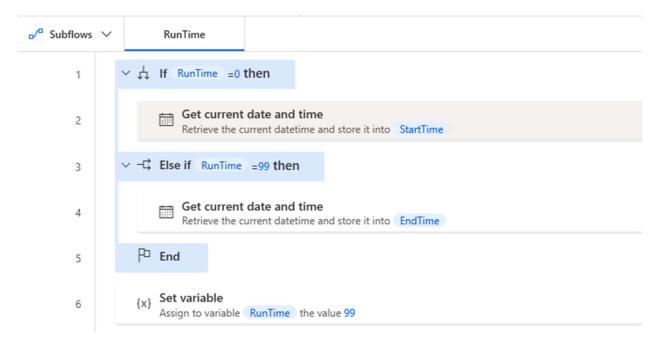
1. RunFlow

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.



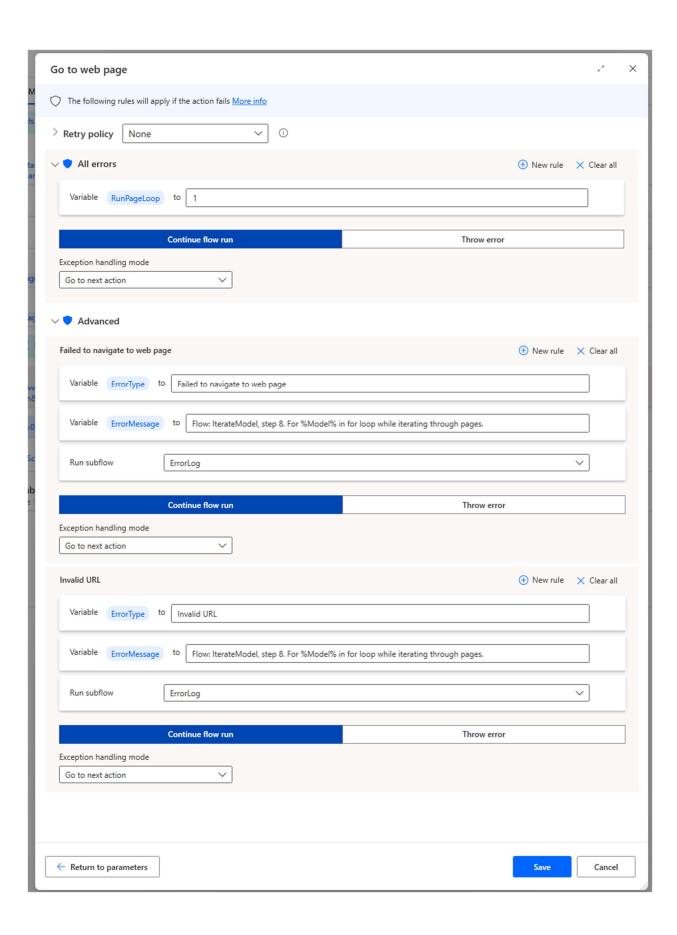
2. RunTime

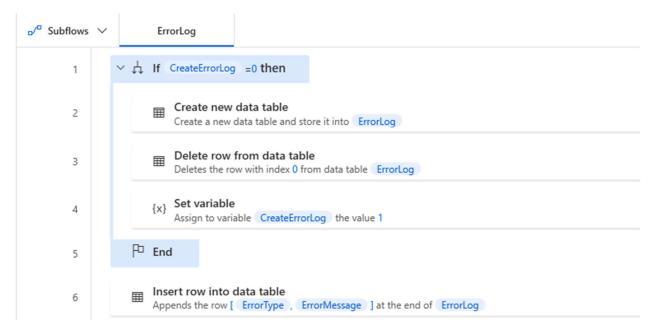
a. A simple sub flow to capture the start time at the beginning when the RunTime variable is set to 0. It then changes to 99 so once the flow is triggered again at the end it sets the time to EndTime instead of StartTime.



3. ErrorLog

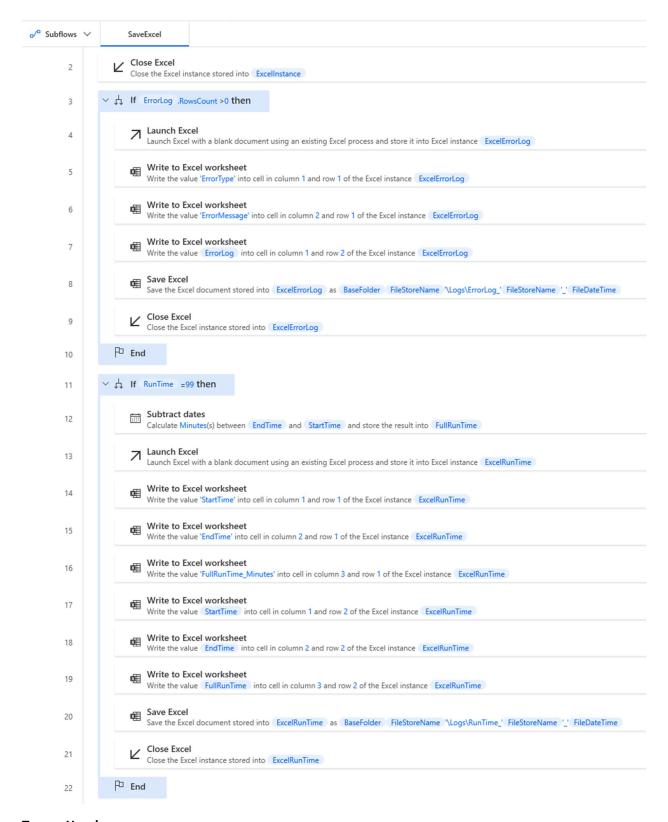
- a. An error log is created to track issues that happen within the flow. This allows you to continue running the rest of the flow if it is not a critical error. The error log is later written to an Excel file with all errors incurred in the flow.
- b. The errors are produced by setting error rules within 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.



Tampa Honda

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.												