⚠ Events are currently unavailable while we investigate this mornings outage (26.06.2025).

GCS: GETCHA Scraper

작성자: Hyewon Shin (FG-AP-52), 마지막 업데이트: 25 6월 2025 • 3분 읽기

Target date	2025. 5. 31
Requestor	@ 알 수 없는 사용자 (qte8936)
User Story	KRITMNGT-4839 - GCS: Getcha Scraper IN PROGRESS
Document status	DRAFT
Document owner	@ Hyewon Shin (FG-AP-52)
Developers	@ Hyewon Shin (FG-AP-52)
Business User	@ Gayoung Ryu (C3-KR-V-1)

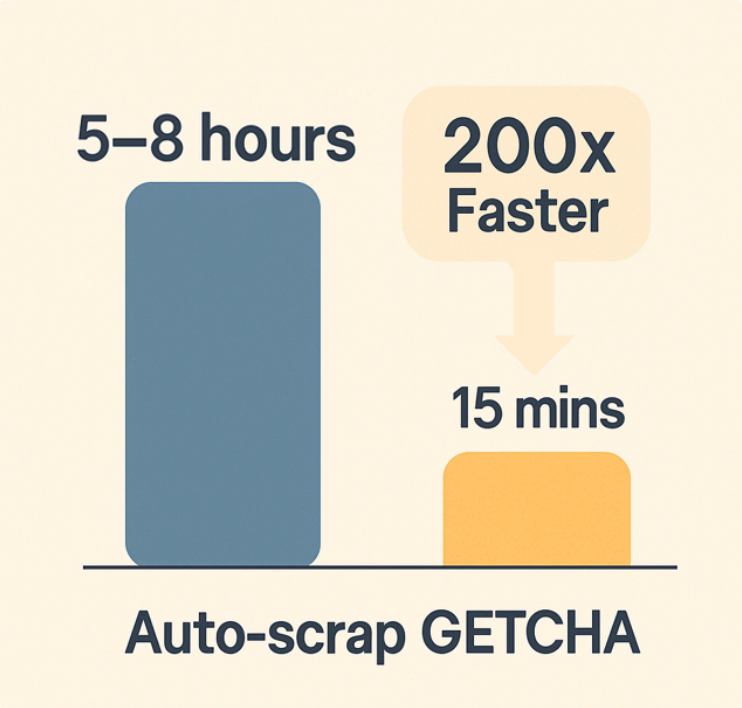
Webpage

https://gcs.bmwkr.ap-northeast-2.aws.cloud.bmw/
Can be accessed within BMW Network.

Goals

- This system is a Python-based web scraper that automatically collects discount information for imported car brands from the GETCHA website (mobile) and saves the data as an Excel file.
- Thereby replacing manul data gathering and ennhance business efficiency.

Background and Strategic Fit



- Existing GETCHA scraping tool has been outdated due to GETCHA website renewal.

- Autmoation system for daily monitoring of BMW's discount rates to ensure price stability.
- Previous human scraping was done **twice a month**, at the beginning and end of each month. Each time it took **at least half a day, often a full day,** making it quite a draining job with a possibility of human error.
- Daily monitoring of competitor's (MB) discount rates to track competitor's price strategy/competitiveness

Key Features

- Brand Series Exploration - Automatically browses each car brand’s model series with discounts
- Model Discount Collection - Retrieves MSRP, cash discounts, and finance discounts
- Fuel Type and Year Detection - Extracts fuel type (P, D, BEV, PHEV) and model year
- Duplicate Data Prevention - Avoids collecting the same series data multiple times per day
- Excel File Output - Saves data daily as `car_data_YYYYMMDD.xlsx`
- Logging & Exception Handling - Writes logs and warnings to `scraping.log` for tracking

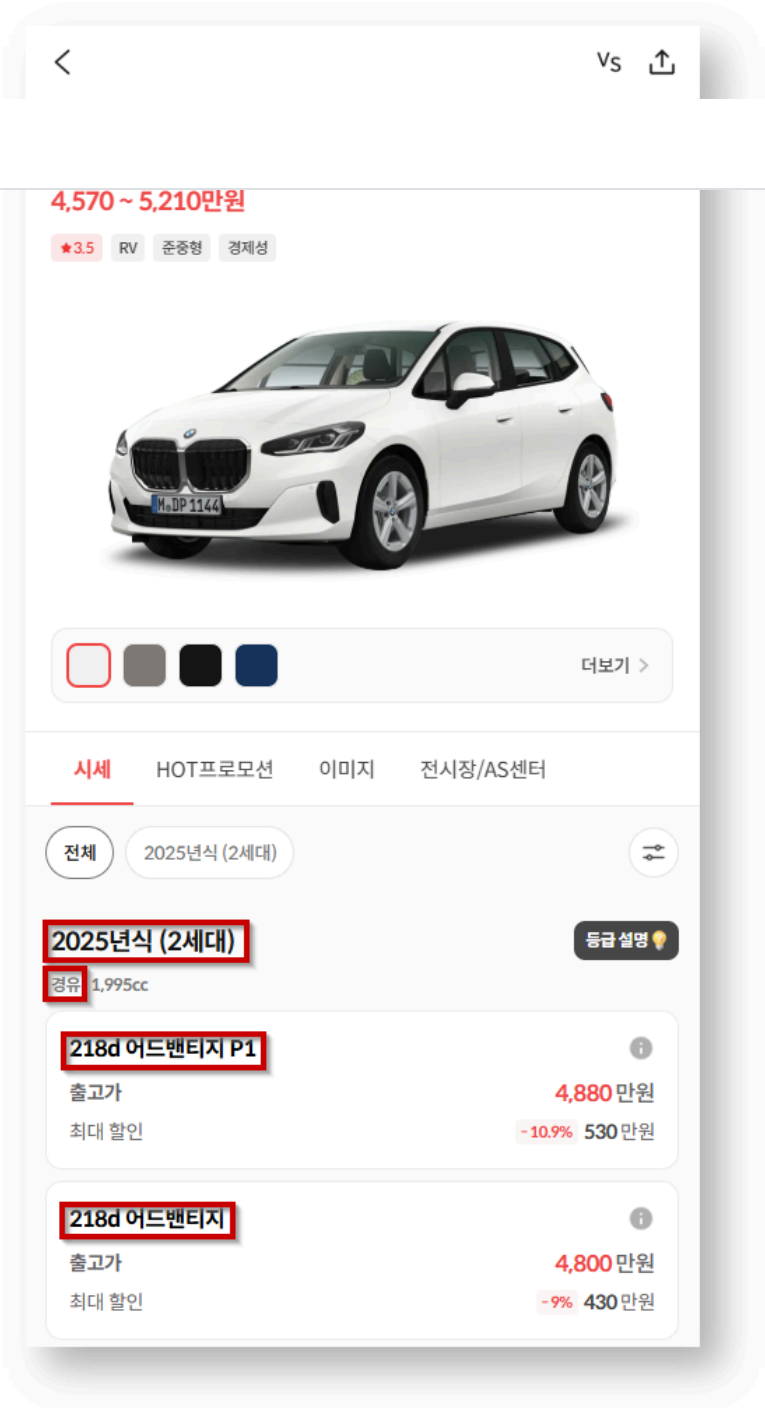
Expected Benefits

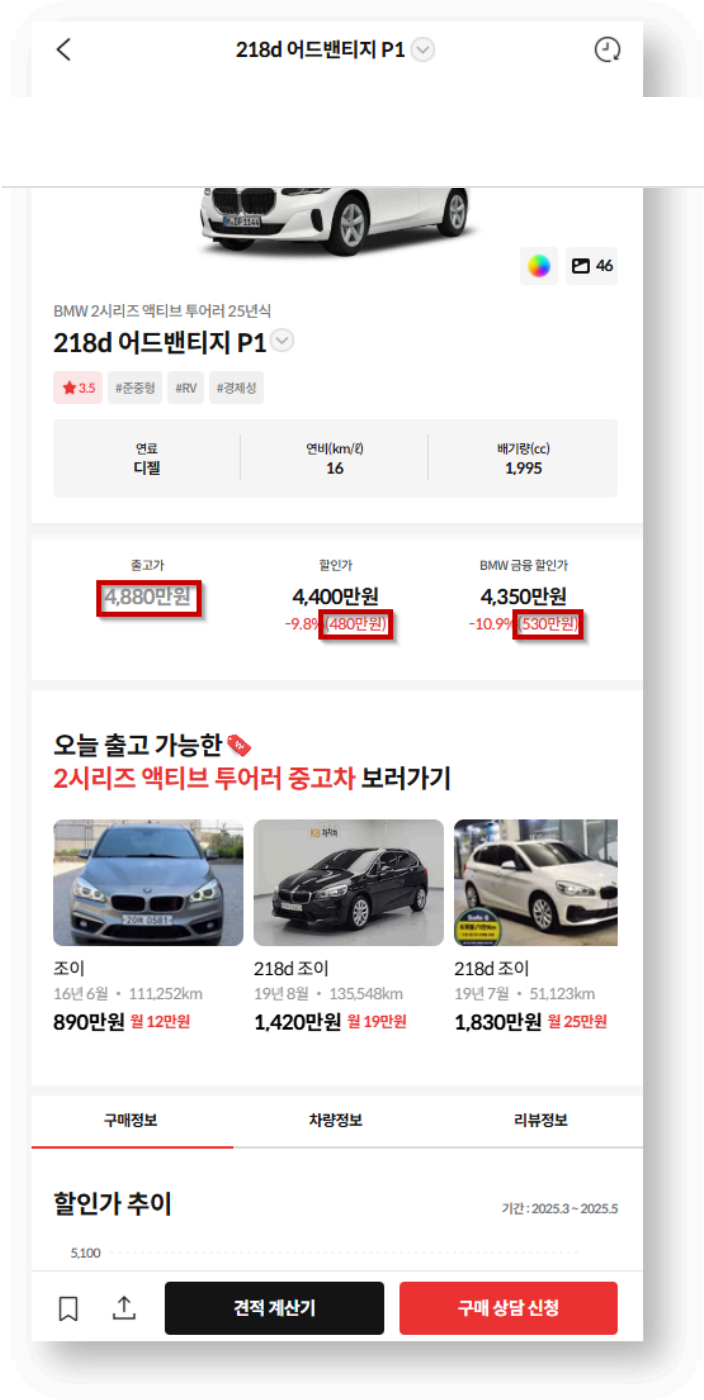
- Daily automated data collection
- Minimized human errors and data omissions
- Real-time competitive insight for Marketing/Strategy teams
- Stable, scalable system without manual intervention

Scope of Data Collection

- Brands Covered: BMW, Mini, Mercedes-Benz, Audi, Volkswagen
- Data Fields: Scraped Date, Model Name, Model Series, Model Year, Fuel Type, MSRP (Manufacturer’s Suggested Retail Price), Cash Discount, Finance Discount

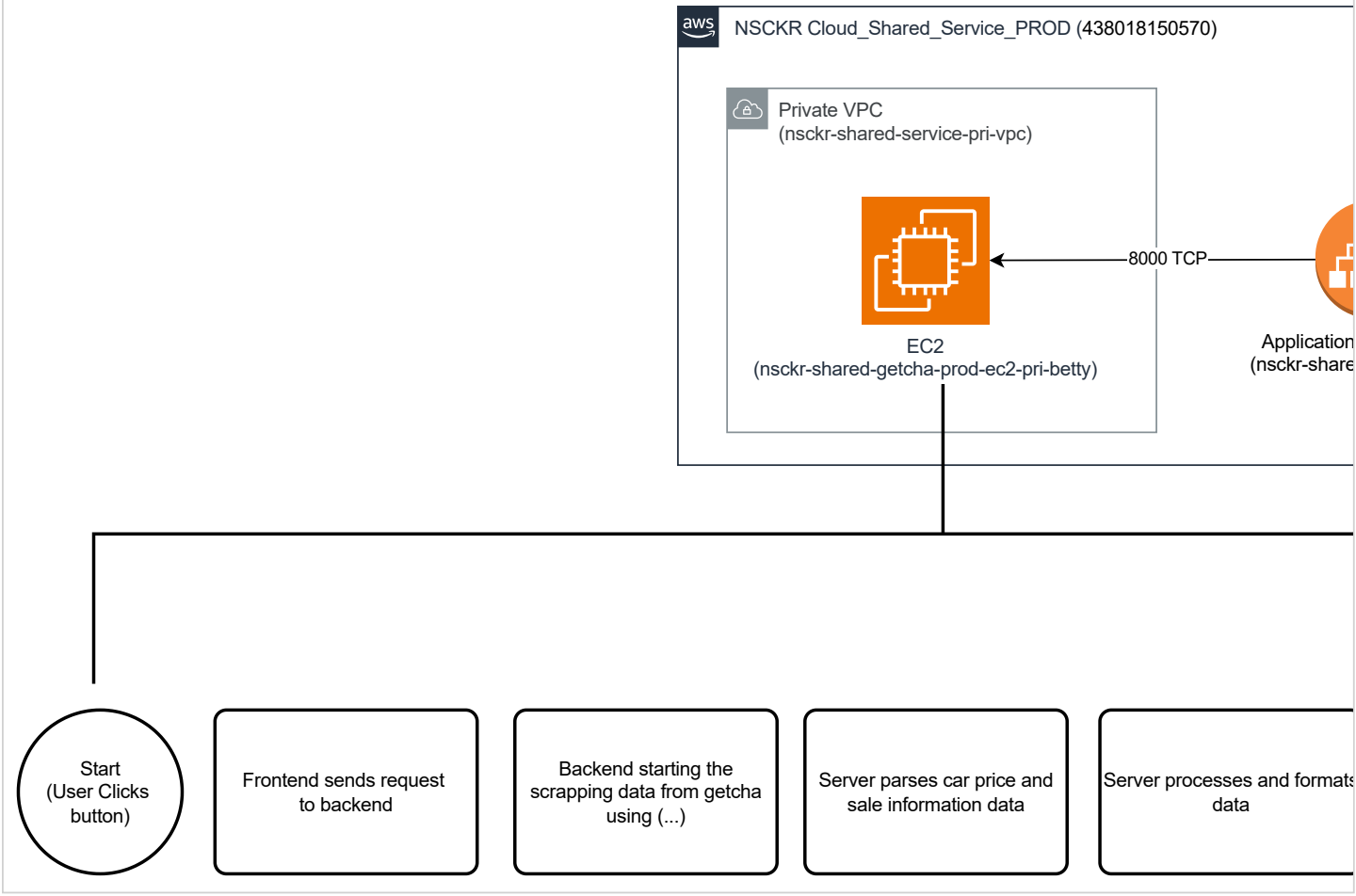
BMW
5시리즈 25년식
520i M 스포츠
10.3% 6,560 만원
1
5시리즈 25년식
520i
10.9% 6,120 만원
2
5시리즈 25년
520i M 스포츠 P
10.9% 6,510
실시간 실거래가
가나다순
모델
출시상태
차종
연료
1시리즈
4,730~6,260만원
#준중형 #해치백 #컴팩트
2시리즈 그란쿠페
4,900~6,430만원
#준중형 #세단 #컴팩트
2시리즈 액티브 투어러
4,570~5,210만원
#준중형 #RV #경제성
2시리즈 쿠페
7,310~7,470만원
#준중형 #쿠페 #스포티
3시리즈
5,790~8,390만원
#중형 #세단 #스포티

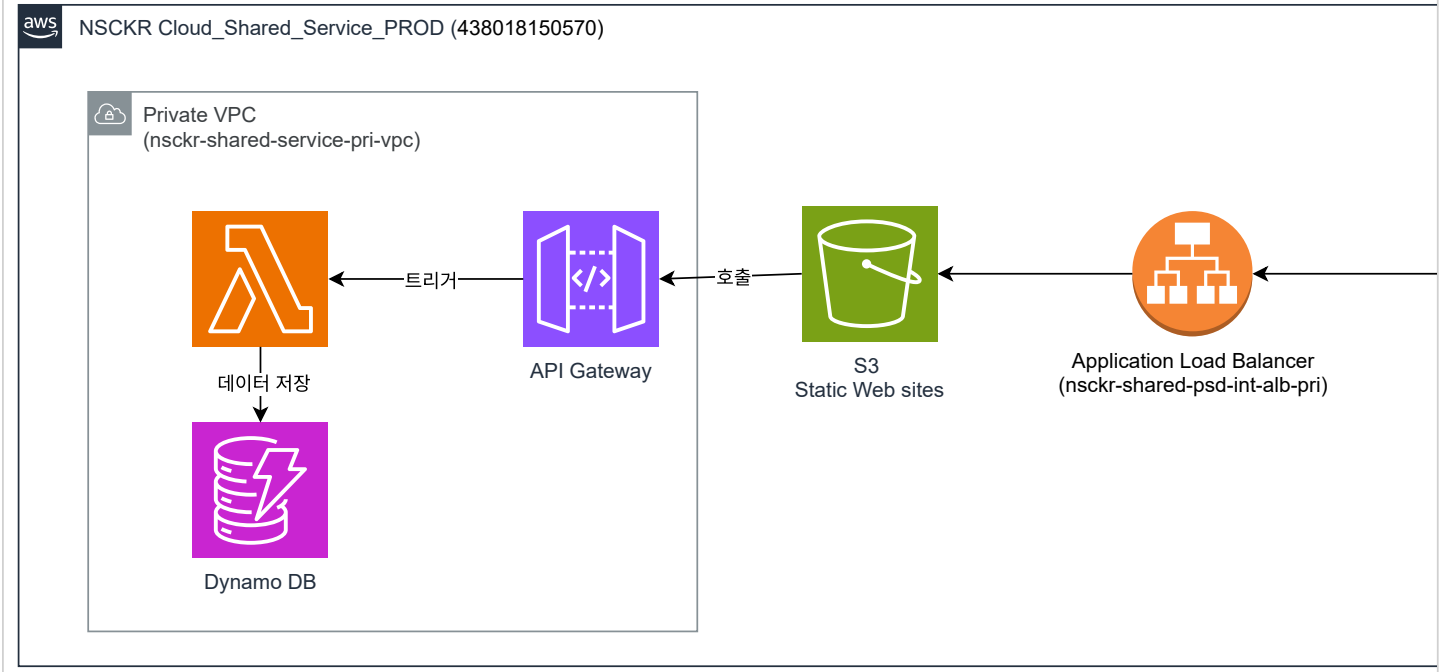

Vs
4,570 ~ 5,210만원
★3.5
RV
준중형
경제성
더보기
시세
HOT프로모션
이미지
전시장/AS센터
전체
2025년식 (2세대)
2025년식 (2세대)
등급 설명
경유 1,995cc
218d 어드밴티지 P1
출고가
4,880 만원
최대 할인
-10.9% 530 만원
218d 어드밴티지
출고가
4,800 만원
최대 할인
-9% 430 만원



Output Data Example

	A	B	C	D	E	F	G	H	I	J	K
1	Year	Month	Date	Brand	MY	Series	Fuel Type	Model (adjusted)	MSRP	Cash_off	Finance_off
2	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리 P1	5,210	430	430
3	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리	5,130	500	500
7	2025	5	7	04_Mini	25	5도어 해치	P	쿠퍼 C 클래식	4,310	43	129
8	2025	5	7	04_Mini	25	5도어 해치	P	쿠퍼 S 페이버드	4,910	49	149
9	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 어드밴티지 P1	4,880	480	530
10	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 어드밴티지	4,800	430	430
11	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리 P1	5,210	430	430
12	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리	5,130	500	500
13	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 어드밴티지 P1	4,650	550	630
14	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 어드밴티지	4,570	700	700
15	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 럭셔리 P1-0	4,950	450	450
16	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 럭셔리 P1	4,950	430	430
17	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 럭셔리	4,870	500	600
18	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 어드밴티지 P1	4,880	480	530
19	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 어드밴티지	4,800	430	430
20	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리 P1	5,210	430	430
21	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리	5,130	500	500
22	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 어드밴티지 P1	4,650	550	630
23	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 어드밴티지	4,570	700	700
24	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 럭셔리 P1-0	4,950	450	450
25	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 럭셔리 P1	4,950	430	430
26	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 럭셔리	4,870	500	600
27	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 어드밴티지 P1	4,880	480	530
28	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 어드밴티지	4,800	430	430
29	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리 P1	5,210	430	430
30	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	D	218d 럭셔리	5,130	500	500
31	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 어드밴티지 P1	4,650	550	630
32	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 어드밴티지	4,570	700	700
33	2025	5	7	01_BMW	25	2시리즈 액티브 투어러	P	220i 럭셔리 P1-0	4,950	450	450


NSCKR Cloud_Shared_Service_PROD (438018150570)
Private VPC
(nsckr-shared-service-pri-vpc)
8000 TCP
EC2
(nsckr-shared-getcha-prod-ec2-pri-betty)
Application
(nsckr-share
Start
(User Clicks
button)
Frontend sends request
to backend
Backend starting the
scrapping data from getcha
using (...)
Server parses car price and
sale information data
Server processes and formats
data



Tech Stack

- Python 3.11
- Playwright (async browser automation)
- BeautifulSoup (HTML parsing)
- pandas (data handling and Excel export)
- openpyxl (Excel writing engine)
- Flask

Future Improvements

Deploy the system to a shared server (NSCKR Cloud Shared Service Sandbox)
Implement data integrity checks by cross-validating with GETCHA website. (Scraping code currently works on GETCHA Mobile)
Lambda Migration
Integrate with MCP for automated code maintenance and lifecycle manangement.

Requirements

#	Title	Importance	Done
1	Web scaping is designed to ensure full compliance with applicable local legal and regulatory frameworks	Must Have	O
2	End-User Training		O
3	Regularly accomodate to changes in website structure		

User Requirements and Interaction

- NSCKR Github Access - Licence owned by FG-AP-52

Questions

Below is a list of questions to be addressed as a result of this requirements document:

Question	Outcome

Not Doing

- Maintenance

레이블 없음