

현대자동차 국내사업본부  
Data 분석 인턴 공모전

★  
**해 카 톤**  
HACARTHON

**과제 설명서**

Welcome to **HACARTHON** !

현대자동차 국내사업본부 Data 분석 인턴 공모전

## 『해카톤』 참가를 환영합니다.

### □ 현대자동차 국내사업본부 『해카톤』 혜택

현대자동차 국내사업본부 『해카톤』은 Data 우수 인재 채용을 위한 인턴 공모전입니다.  
『해카톤』 결과에 따라 수상자/우수자가 선정되고, 채용전환형 인턴프로그램을 통해  
현대자동차 국내사업본부 신입사원으로 입사할 수 있는 기회를 잡으세요!

1등	2등	우수자
1명	2명	00명
채용전환형 인턴 프로그램 수행 (필수 사항)		
상장 + 상금 300만 (인턴 급여에 포함)	상장 + 상금 100만 (인턴 급여에 포함)	상장

\* 『해카톤』은 우수 인재 채용 목적의 공모전으로 수상자/우수자로 선정되신 분들은 반드시  
인턴 프로그램에 참여해야 하며, 인턴 미참여 시 수상 내역(상금포함)이 취소됩니다.

### □ '21년 『해카톤』 진행 일정

4/13(화)	6/14(월) 오전 9시	6월말	7월중	8월	9월~
참가확정 및 과제안내	결과물 제출 마감	1차 심사 (서류)	2차 심사 (PT)	채용전환형 인턴프로그램 (1개월)	평가/입사

\* 심사 및 인턴 프로그램 세부 일정은 대상자 한 별도 안내 예정이며, 당사 사정에 따라 세부  
일정은 변경 될 수 있습니다.

## □ 『해카톤』 과제 선정 배경 및 재구매 유형 구분

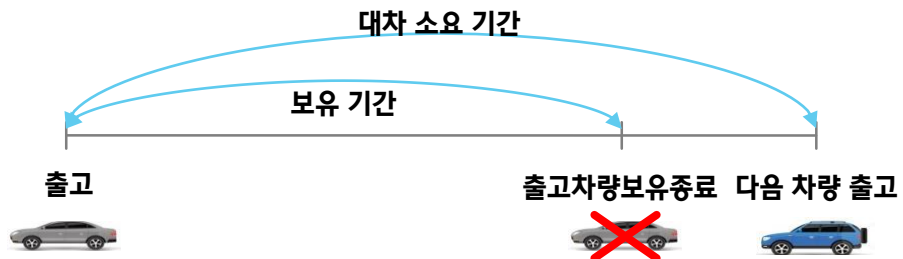
현대자동차 국내사업본부는 현대자동차 구매 고객 대상으로 재구매 유도를 위한 개인화 마케팅을 진행하고 있으며, 고객의 재구매 경향 분석을 기반으로 고객에게 최적합 차량을 제공하기 위해 노력하고 있습니다.

자동차 재구매 유형은 ‘대차’와 ‘추가구매’로 분류할 수 있으며, ‘구매이력’과 ‘차량 보유정보’ 등을 활용하여 재구매 유형 별 구매 시점을 예측하는 것이 재구매 분석의 핵심으로 인지하고 있습니다.

이번 인턴 공모전 『해카톤』과제 주제는 가상 고객 구매 Data 기반으로 ‘대차’와 ‘추가구매’ 유형을 구분하고 EDA 및 모델링을 통해 ‘대차’ 또는 ‘추가구매’ 가능성을 판단하는 것으로, 『해카톤』참가자 분들의 참신하고 도전적인 시도를 기대합니다.

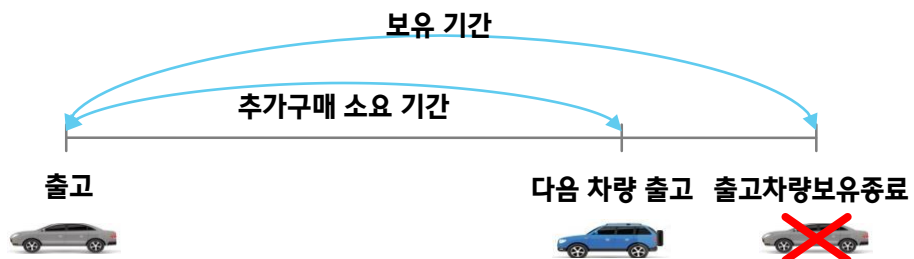
### 대차

기존 보유하던 차량을 처분하고, 신차를 출고하는 경우로 대차 소요 기간과 보유 기간이 유사하며, 보유 기간이 대차 소요 기간 보다 길 수도 있음

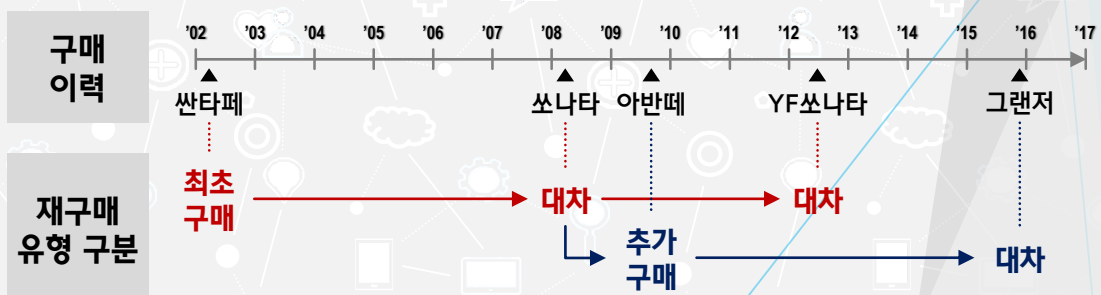


### 추가 구매

기존 보유하던 차량을 처분하지 않고, 추가로 신차를 출고하는 경우로 보유 기간과 추가구매 소요 기간이 일정 수준의 차이를 보임



### ※ 구매 이력에 따른 재구매 유형 분류 예시



## □ 『해카톤』 과제 설명 및 결과 제출물

### 과제 1

주어진 가상 고객 구매 Data를 활용하여 ‘대차’와 ‘추가구매’를 분리/추정하는 로직을 개발하고, ‘대차’, ‘추가구매’ 구분 결과에 대해 각각 EDA를 수행하여 고객과 차량 관점에서 특징/차이점을 도출하시오

#### 결과 제출물

- ① 재구매 유형 분리/추정 로직 설명서 및 코딩 소스
- ② EDA 결과서 및 코딩 소스 (Data 분석 및 인사이트 도출)

#### 유의 사항

- \* ‘대차’와 ‘추가구매’의 분리/추정 방식은 자율이며, 논리전개 과정과 근거를 반드시 기재해 주시기 바랍니다.
- ex. 기존 보유 차량의 처분 시점 전후 3개월 내 신차 구매를 ‘대차’로 가정

### 과제 2

고객 별 6개월 내, ‘대차’ 또는 ‘추가구매’ 가능성(Score)에 대한 예측모델을 개발하시오

#### 결과 제출물

- ① 예측모델 결과서 및 코딩 소스

#### 유의 사항

- \* ‘차량 별’이 아닌 ‘고객 별’ 예측모델이며, ‘대차’ 또는 ‘추가구매’ 중 1개를 선택하여 모델링할 수 있습니다.
- \* 과제1의 대차/추가구매 분리 알고리즘을 활용한 종속변수 생성 및 주어진 Data를 활용한 고객 별 독립변수를 생성하여 예측모델을 구현합니다.
- \* 예측모델 결과서는 ‘문제 정의, 샘플링 정의, 독립/종속 변수 정의, 모델링, 평가지표 선정, 모델링 결과 및 검증, 활용 방안’을 반드시 포함해야 합니다.

#### ※ 과제 수행 및 결과물 제출 주의 사항

- 1) 주어진 Data set 외 참가자 의도에 따라 공공Data, 외부 Data를 결합하여 사용할 수 있으며, 해당 Data와 출처를 함께 제출해야 합니다.
- 2) 분석 Tool은 반드시 Python을 활용하고, 코딩 설명 주석 처리 및 결과 재현은 필수입니다.
  - 주어진 Python 분석 환경에서 Data Load, 가공, EDA, 모델링을 수행하여야 합니다.
  - 코딩 소스는 .ipynb (&.html) 로 제출하며, 주어진 환경에서 재현되어야 합니다.
  - Python 이외 분석 Tool (R, SAS 등) 활용 시 심사에서 제외 됩니다.
- 3) 설명서/결과서는 파워포인트로 제출하며, 과제 1,2 각 10장을 넘지 않습니다.

※ 과제 제출처 : 현대자동차 국내사업본부 이메일 KoreaBizHR@hyundai.com

## □ 『해카톤』 과제 수행을 위한 Data set

① 고객 정보, ② 차량 정보, ③ 고객 접촉 이력 정보 로 3개로 구성되어 있습니다.  
각 Data 별 아래 내용을 확인하시어, '21년 『해카톤』 과제를 수행해 주시기 바랍니다.

★ 본 Data set 은 현대자동차 소유로 '21년 『해카톤』 참가자에게만 제공되며,  
타인 또는 타기관 등 어떤 누구에게도 Data 공유를 금합니다. 아울러, 보안 규칙에  
따라 Data 추적이 가능하며, 유출 시 참가자에게 법적 책임을 물을 수 있습니다.

### ① 고객 정보

#### ■ 개요

- 첨부 파일명 : '21년 해카톤\_고객정보.csv
- 구분자(Delimiter) : 콤마(,)
- Data 건수 : 1,096,206 건

#### ■ 내용

- 차량 구매 이력이 있는 고객의 Demo 정보 수록
- 주소 Data는 고객 별로 Data가 없을 수 있음

No	변수명 (영문)	변수명 (한글)	타입	길이	PK 여부	코멘트
1	CUS_ID	고객ID	문자	23	Y	
2	PSN_BIZR_YN	개인사업자여부	문자	1		고객의 사업자 여부 표기 'Y' : 사업자등록이 된 고객
3	SEX_SCN_NM	성별	문자	10		
4	TYMD	생년월	문자	8		생년월까지만 표기, 일자는 '00'으로 일괄 표기
5	CUS_ADM_TRY_NM	주소_행정시도명	문자	100		고객 주소 표기
6	CUS_N_ADMZ_NM	주소_시군구명	문자	100		고객 주소 표기
7	CUS_ADMB_NM	주소_행정동명	문자	60		고객 주소 표기
8	CLB_HOUS_PYG_NM	주택 평형	문자	20		고객 거주 주택의 평형
9	REAI_BZTC_AVG_PCE	주택 평균가격	숫자	8		고객 거주 주택의 가격

## □ 『해카톤』 과제 수행을 위한 Data set

### ② 차량 정보

#### ■ 개요

- 첨부 파일명 : '21년 해카톤\_차량정보.csv
- 구분자(Delimiter) : 콤마(,)
- Data 건수 : 1,835,830 건

#### ■ 내용

- 판매 된 신차 차량 정보 수록 (중고차 판매 이력 없음)
- 고객과 차량은 1:N 관계로 한 명의 고객이 다수의 차량 구매 이력 존재
- 출고 일자와 보유시작일이 언제나 같지 않음
- 차명, 차량등급, 엔진타입, 트림에 대한 설명은 별도 제공 하지 않으며, 참가자 역량으로 파악할 것

No	변수명 (영문)	변수명 (한글)	타입	길이	PK 여부	코멘트
1	CAR_ID	차량ID	문자	26	Y	
2	CUS_ID	고객ID	문자	23		
3	WHOT_DT	출고일자	문자	8		
4	CAR_HLDG_STRT_DT	보유시작일자	문자	8		차량 보유 시작일
5	CAR_HLDG_FNH_DT	보유종료일자	문자	8		차량 보유 종료(처분)일 ※ Null인 경우 미처분상태
6	CAR_NM	차명	문자	100		
7	CAR_CGRD_NM_1	차량등급명1	문자	100		
8	CAR_CGRD_NM_2	차량등급명2	문자	100		
9	CAR_ENG_NM	엔진타입명	문자	100		
10	CAR_TRIM_NM	트림명	문자	100		



## □ 『해카톤』 과제 수행을 위한 Data set

### ③ 고객 접촉 이력 정보

#### ■ 개요

- 첨부 파일명 : '21년 해카톤\_접촉정보.csv
- 구분자(Delimiter) : 콤마(,)
- Data 건수 : 17,734,097 건

#### ■ 내용

- 구매 이력이 있는 고객이 현대자동차에 자발적(Inbound)으로 접촉한 이력 정보
- 출고, 자동차 보유 전후 이력 모두 포함된 자발적(Inbound) 접촉 이력
- 접촉채널, 접촉업무에 대한 설명은 별도로 제공하지 않으며, 참가자 역량으로 파악할 것

No	변수명 (영문)	변수명 (한글)	타입	길이	PK 여부	코멘트
1	CNTC_SN	접촉일련번호	숫자	8	Y	
2	CUS_ID	고객ID	문자	23		
3	CNTC_DT	접촉일자	문자	8		고객이 자발적으로 현대자동차에 접촉한 일자
4	CNTC_CHAN_NM	접촉채널명	문자	256		고객이 자발적으로 현대자동차에 접촉한 채널
5	CNTC_AFFR_SCN_NM	접촉업무명	문자	256		고객이 자발적으로 현대자동차에 접촉한 업무명

## □ 『해카톤』 과제 수행을 위한 Python 분석 환경 설정 및 제출 파일

### ★ 개발 환경

- python==3.6.10
  - \* 리눅스 환경 분석 시 Ubuntu 18.04.5 사용 권장
- 제공된 패키지 사용 권장 (requirements.txt)
  - \* 추가 패키지 사용 시 requirements.txt 를 수정/추가하여 제출

### ★ 제출 파일

- .ipynb 필수 제출 (html 파일 제출 포함) (.py파일 추가 제출 가능)
  - \* 파일명 예시 : {이름}\_{생년월일}\_notebook.ipynb (모든 파일 적용)
- 모든 파일은 한 디렉토리 안에서 압축하여 제출 (원본데이터 포함)
  - \* ex. 김현대\_950101.zip : .ipynb, .html, 원본데이터.csv, src.py(선택), requirements.txt(선택), 설명서/결과서(PPT)
  - \* 파일 입출력 시 상대 경로 사용 필수

## □ 『해카톤』 평가/심사 기준

구분		문제 접근	Data 이해/가공	모델링	리포팅	계
과제 1	재구매 유형 분리 로직 개발	10	20	5	5	40
	EDA	5	5	-	10	20
과제 2	재구매 예측모델 개발	5	10	20	5	40

- ① 상기 평가 점수표 기준으로 1,2차 심사가 진행됩니다.
  - 1차 심사 : 제출한 결과물 기반 별도 평가자 집단이 평가
  - 2차 심사 : 제출한 결과물 프레젠테이션 실시 및 별도 평가자 집단이 평가 (프레젠테이션 방식 추후 안내 예정)
- ② 결과서, 코딩 소스 등 최종 제출물 구성 기준을 충족하지 못하는 경우, 평가/심사에서 자동 제외 됩니다.
- ③ 모든 Data 분석 사항은 주어진 분석 환경에서 재현이 되어야 하며, 주어진 분석 환경에서 분석 재현이 되지 않을 경우 평가/심사에서 불이익을 받을 수 있습니다.
- ④ 주어진 과제를 모두 완료 하지 않아도 분석 접근의 창의성, 독창성 등을 고려하여 부분 점수가 적용됩니다.
- ⑤ 공공 Data, 외부 Data 활용은 자율이나, 불필요한 추가 Data 적용은 감점 요인이 될 수 있습니다.
- ⑥ 제출된 결과물이 심사 기준에 부합하지 않을 경우, 수상작을 선정하지 않을 수 있으며, 제출된 자료에 대한 권리는 현대자동차가 소유합니다.
- ⑦ 현대자동차는 표절 및 저작권 관련 법규/규정을 준수합니다. 표절 및 저작권 관련 문제의 소지가 있는 결과물은 심사에서 제외되며, 수상 후에도 수상이 취소되고 모든 책임은 참가자에게 있습니다.



## □ 『해카톤』 FAQ

no	FAQ
1	<p>Q) 결과물이 Python코드와 html만 제출 가능한가요?</p> <p>A) 분석 환경 설명서에 명시된 코드 소스만 심사 대상이 됩니다. Python 코드 이외 코드( R, SAS, SQL 등)의 코드 제출 시 심사 대상에서 제외 됩니다.</p>
2	<p>Q) 과제1, 과제2를 모두 수행하여야 하나요?</p> <p>A) 과제1, 과제2를 모두 완성하지 않아도 과제 수행 상의 논리성과 창의성에 평가 부분 점수가 부여됩니다. 지원자의 역량이 충분히 표현될 수 있도록 수행 결과를 최대한 정리하여 제출하는 것을 권고 드립니다.</p>
3	<p>Q) 제공된 과제 설명 및 Data set에 대한 개별적인 추가 질문이 가능한가요?</p> <p>A) 과제 및 Data에 관한 추가 설명은 제공하지 않습니다. 본 설명서에 표현되지 않은 부분은 지원자의 자율적인 판단에 따릅니다.</p>
4	<p>Q) 인턴 공모전 『해카톤』과 타 채용 공고에 중복 지원한 경우, 어떻게 진행되나요?</p> <p>A) 2차 심사 대상자 분들은 『해카톤』 인턴 또는 타 공고 채용 전형 중 1개 전형을 선택해야 하며, 세부 방식은 1차 심사 합격자 대상 별도 안내드릴 예정입니다.</p>
5	<p>Q) 향후 진행할 인턴 활동은 어떤 내용인가요?</p> <p>A) Data 분석 유관 부서에서 1개월 간 인턴 프로그램을 수행하게 되며, 공모전 결과와는 별개의 평가를 진행하여 최종 신입사원 입사 여부를 판단합니다.</p>
6	<p>Q) 공모전 상금은 어떻게 지급되나요?</p> <p>A) 수상자 상금은 인턴 급여에 포함되어 지급되며, 각종 제세공과금 등은 본인 부담입니다. 아울러, 인턴 프로그램 미참여 시 수상 내역은 취소됩니다.</p>
7	<p>Q) '22년 2월 졸업예정자의 경우, 신입사원 최종 합격 시 언제 입사하나요?</p> <p>A) '22년 2월 졸업예정자 분들은 '22년 1월 입사 예정이며, 기졸업자/'21년 8월 졸업예정자 분들은 9월 입사 예정입니다.</p>

※ 각종 문의 : 현대자동차 국내사업본부 이메일 [KoreaBizHR@hyundai.com](mailto:KoreaBizHR@hyundai.com)