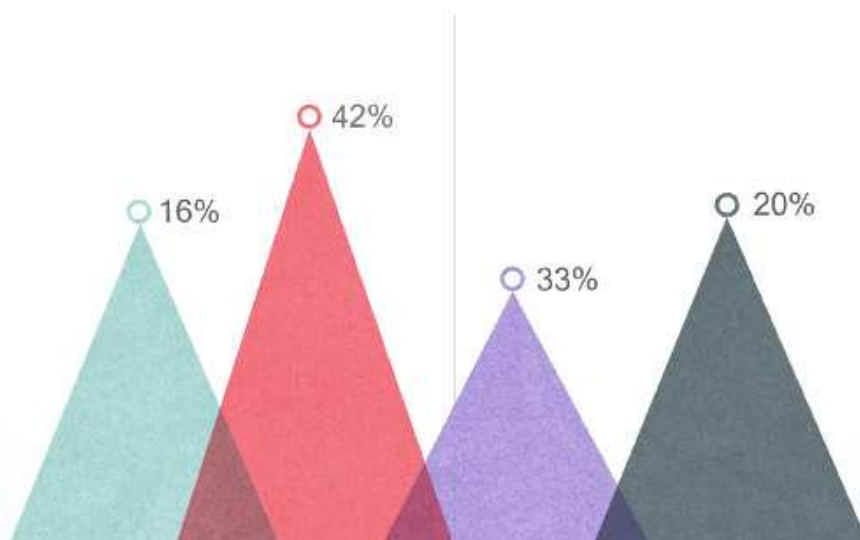


과제

[데이터 사이언스 과정] 프로젝트 과제



공지 및 프로젝트 소개

1. 프로젝트 과제는 **데이터사이언스 자격증 발급과 관련된 중요한 종합과제**입니다.
2. 본 과제는 실제 와이즈인컴퍼니에서 수행한 분석 데이터를 기반으로 합니다. 따라서 **학습 목적 이외에 데이터의 외부 유출을 엄격히 금합니다.**
3. 모듈1 데이터분석과 시각화 종합과제는 중고자동차 매매가격 예측 제공 서비스 사업으로 현재 와이즈인컴퍼니와 오토업(autoup)에서 공동 사업운영하고 있는 중고차 매매가격 데이터를 기반으로 중고차를 거래할 때 얼마 정도가 적합한지를 예측, 제공하는 정보제공 사업입니다.
4. 정답이 정해져 있는 것이 아니기때문에 세부 과제별로 최선의 방법을 모색하고, 학습자의 판단에 따라 분석결과를 정리해야 합니다. 평가 역시 **정답이 아닌 분석 로직(logic)과 분석 효율성을 중심으로** 하며, 학습 이외의 내용이 과제로 제출될 수 있습니다. 학습 이외의 과제는 다양한 정보(구글링 등)를 통합하여 해결하시기 바랍니다.
5. 본 과제 제출마감은 **2019년 3월 24일(일요일)**까지입니다. 과제물은 wise@wiseinc.co.kr로 제출 바랍니다.
 - 제출 과제물: 1) 결과물 정리 ppt 2) python 코드문서 3) 분석에 적용된 데이터파일
6. 본 과제의 풀이는 화상강의가 아닌 별도의 영상강의를 추후 제공할 예정입니다 (홈페이지 강좌 업로드)

1 과제

1. 데이터구조 및 과제1

data 구조

1	차량번호	제조사	모델	세부모델	등급	세부등급	차종	상세차종	형식년도	주행거리	연료	사고유무	변속기	쎌루프	내비게이션	서라운드뷰	판매가격(만원)
2	13너5836	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	디렉스	승용	경차	2013	90,301	휘발유	무사고	4단 자동				1
3	25거2493	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	트렌디	승용	경차	2015	43,736	휘발유	무사고	4단 자동				1
4	27오9027	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	디렉스	승용	경차	2013	204,000	휘발유	무사고	4단 자동				1
5	30우2260	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	트렌디	승용	경차	2014	55,358	휘발유	무사고	4단 자동				1
6	34오5344	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2012	134,000	휘발유	무사고	4단 자동				1
7	35로8463	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	디렉스	승용	경차	2012	165,143	휘발유	무사고	4단 자동				3
8	33오7671	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	디렉스 스페셜	승용	경차	2012	105,951	휘발유	무사고	4단 자동				6
9	51버4650	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	디렉스 스페셜	승용	경차	2012	48,000	휘발유	무사고	4단 자동				7
10	14마8522	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	디렉스 스페셜	승용	경차	2012	59,815	휘발유	무사고	4단 자동				10
11	14모8229	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2015	22,722	휘발유	무사고	4단 자동				10
12	16조2982	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2015	15,000	휘발유	무사고	4단 자동				10
13	16버8292	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	디렉스	승용	경차	2012	80,000	휘발유	무사고	4단 자동				10
14	20나8682	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2013	41,995	휘발유	무사고	4단 자동		내비게이션		10
15	22러2241	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2012	70,000	휘발유	무사고	4단 자동				10
16	23모1274	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2013	100,549	휘발유	무사고	4단 자동		내비게이션		10
17	33모2139	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	트렌디	승용	경차	2014	42,413	휘발유	무사고	CVT				10
18	35나4207	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2015	80,212	휘발유	무사고	4단 자동	쎌루프			10
19	38수3287	기아	모닝	올 뉴 모닝 (TA 11.1.24~15.2)	1.0 카파	럭셔리	승용	경차	2012	108,000	휘발유	무사고	4단 자동	쎌루프			10

분석 목적

본 분석의 목적은 중고 자동차의 판매가격(만원)을 예측하기 위한 최적의 회귀모델을 구축하여, 새로운 중고자동차가 거래될 때 (기존의 거래가격을 바탕으로) 합리적인 중고거래가를 예측 및 책정하기 위함이다.

1

과제

2. 과제2

과제1(배점 10점)

우선 판매가격에 미치는 영향변수가 될 수 있는 예측자(predictor) 혹은 독립변수(independent variable)을 선정하여 모형을 도식화하시오.

과제2(배점 20점)

각 변수별 데이터를 탐색하여 결측치 제거, 이상치 제거, 필요한 데이터 변환(정규화, 범주화 등)을 수행하시오

과제3(배점 20점)

예측변수와 판매가격 간에 상관성이 있는지 분석을 수행한 후 결과정리 및 상관성이 유의한 변수를 선정하시오

1

과제

2. 과제2

과제4(배점 20점)

과제3에서 상관이 있는 예측변수만을 선정하여 판매가격에 미치는 분석을 수행하고 결과를 정리 및 해석하시오.

단, 표준화잔차값이 절대값 2보다 큰 자료는 제거하고 최종 모델을 구성하시오

과제5(배점 20점)

세부등급별로 판매가격 예측모형을 만들되, 세부등급 거래 데이터가 150개 이상인 모델만 분석하고 과제4와 동일하게 결과를 정리 및 해석하시오.

과제6(배점 10점)

임의적으로 새로운 거래데이터를 3개 정도 엑셀파일로 만들어서 이를 붙여와 바로 예측가가 제시될 수 있도록 python 함수를 만들어 적용하시오.

2

결과물 정리

1. 과제1

과제1(배점 10점)

우선 판매가격에 미치는 영향변수가 될 수 있는 예측자(predictor) 혹은 독립변수(independent variable)을 선정하여 모형을 도식화하시오.

모형(도식화) 제시

1 과제

2. 과제2

과제2(배점 20점)

각 변수별 데이터를 탐색하여 결측치 제거, 이상치 제거, 필요한 데이터 변환(정규화, 범주화 등)을 수행하시오

데이터 정제 유형	기존데이터	변환 후 데이터	제거/변환 근거(이유)
결측치 제거	2938개	0000개 (00개 제거)	
이상치 제거-판매가격	0000개	0000개 (00개 제거)	
.....			
범주화-연료	기존범주 4개		
.....			

결과 정리 예시

1 과제

2. 과제2

과제3(배점 20점)

예측변수와 판매가격 간에 상관이 있는지 분석을 수행한 후 결과정리 및 상관이 유의한 변수를 선정하시오

예측변수	분석방법	상관여부	최종 예측변수 선정여부
주행거리			선정/제외
형식년도			
연료			
...			

결과 정리 예시

1 과제

2. 과제2

과제4(배점 20점)

과제3에서 상관이 있는 예측변수만을 선정하여 판매가격에 미치는 분석을 수행하고 결과를 정리 및 해석하시오.

단, 표준화잔차값이 절대값 2보다 큰 자료는 제거하고 최종 모델을 구성하시오.

	계수*				유의확률
	비표준화 계수		표준화 계수		
	B	표준오차	베타	t	
(상수)	-52523.069	3567.185		-14.724	.000
형식년도	26.326	1.771	.227	14.868	.000
주행거리	-.002	.000	-.473	-30.764	.000
연료_LPG	43.372	52.255	.028	.830	.417
연료_LPG겸용	48.620	49.133	.053	.990	.322
연료_휘발유	24.050	47.373	.011	.708	.481
변속기_없음	151.806	23.234	1.130	6.534	.000
변속기_4단자동	168.764	12.821	.225	13.163	.000
변속기_CVT	105.191	31.006	.052	3.393	.001
싼루프_있음	58.308	6.361	.134	9.166	.000
내비게이션_있음	52.571	5.987	.128	8.780	.000

차량정보	상수	형식년도	주행거리	연료	변속기	싼루프	내비게이션	추정가격
		2012	10000	디젤	5단수동	없음	없음	-52523.069
		2012	10000	LPG	4단자동	있음	있음	52968.7659
				LPG겸용	CVT	없음	없음	-18.378175
				휘발유	5단수동	없음	없음	0
				없음	없음	있음	없음	0
				4단자동	CVT	없음	없음	168.764383
				CVT	없음	없음	없음	0
				싼루프	없음	있음	없음	58.3081499
				내비게이션	없음	없음	있음	0
				있음	있음	없음	있음	52.5714834
								추정가격
								706.962

해석:

1 과제

2. 과제2

과제5(배점 20점)

세부등급별로 판매가격 예측모형을 만들되, 세부등급 거래 데이터가 150개 이상인 모델만 분석하고 과제4와 동일하게 결과를 정리 및 해석하시오

	계수*				유의확률
	비표준화 계수		표준화 계수		
	B	표준오차	베타	t	
(상수)	-52523.069	3567.185		-14.724	.000
형식년도	26.326	1.771	.227	14.868	.000
주행거리	-.002	.000	-.473	-30.764	.000
연료_LPG	43.372	52.255	.028	.830	.417
연료_LPG겸용	48.620	49.133	.053	.990	.322
연료_휘발유	24.050	47.373	.011	.708	.481
변속기_없음	151.806	23.234	1.130	6.534	.000
변속기_4단자동	168.764	12.821	1.275	13.163	.000
변속기_CVT	105.191	31.006	.052	3.393	.001
싼루프_있음	58.308	6.361	.134	9.166	.000
내비게이션_있음	52.571	5.987	.128	8.780	.000

차량정보	상수	형식년도	주행거리	연료	변속기	싼루프	내비게이션	추정가격
		2012	10000	디젤	5단수동	없음	없음	-52523.069
		2012	10000	LPG	4단자동	있음	있음	52968.7659
				LPG겸용	CVT	없음	없음	-18.378175
				휘발유	5단수동	없음	없음	0
				없음	없음	있음	없음	0
				4단자동	CVT	있음	있음	168.764383
				CVT	없음	없음	없음	0
				싼루프	없음	있음	없음	58.3081499
				내비게이션	없음	없음	있음	0
				있음	있음	있음	있음	52.5714834
								추정가격
								706.962

해석:

1 과제

2. 과제2

과제6(배점 10점)

임의적으로 새로운 거래데이터를 3개 정도 엑셀파일로 만들어서 이를 불러와 바로 예측가가 제시될 수 있도록 python 함수를 만들어 적용하시오.

임의데이터

1	형식년도	주행거리	연료	사고유무	변속기	썬루프	내비게이션	서라운드뷰	판매가격(만원)
2	2012	65,556	휘발유	무사고	4단 자동	썬루프	내비게이션		680
3	2015	26,978	휘발유	무사고	4단 자동		내비게이션		690
4	2015	30,000	휘발유	무사고	CVT				690



1	형식년도	주행거리	연료	사고유무	변속기	썬루프	내비게이션	서라운드뷰	판매가격(만원)
2	2012	65,556	휘발유	무사고	4단 자동	썬루프	내비게이션		680
3	2015	26,978	휘발유	무사고	4단 자동		내비게이션		690
4	2015	30,000	휘발유	무사고	CVT				690

예측가

추정가격	628.910	오차	613~801
추정가격	570.602	오차	613~801
추정가격	454.457	오차	613~801



수고하셨습니다!

W와이즈인컴퍼니