# PORTFOLIO



유쾌 상쾌 통계~! 김혜인 포트폴리오

# PORTFOLIO

유쾌 상쾌 통계~! 김혜인 포트폴리오



"일단 하고 생각하자"

# Profile

# 김혜인(Kim Hyein)

2000.3.9

Education	2019.3 ~ 성신여자대학교 통계학과 입학 2020.3 ~ 성신여자대학교 정보시스템공학과 복수전공		
Internship	2021.9 ~ 12 양주시청 기획재정부 정보통신과 빅데이터부 인턴		
Certificate	2020.8 컴퓨터 활용능력 1급 2020.8 사회조사분석사 2급 2021.4 데이터 준전문가(ADSP)		
Club Activity	2020.2 ~ 2020.7 Smarter(스포츠 마케터)20기 미디어홍보팀 2021.2 ~ Daja Vu(성신여대 빅데이터 분석 동아리)		



2021년 겨울 인턴이 되어 최대의 능력치를 발휘할 김혜인입니다.

"일단 하고 생각하자"

# 김혜인(Kim Hyein)

2000.3.9

TEL: 010-9652-5997

EMAIL: haein0758@naver.com

# **Education**

2019.2 서울수락고등학교 졸업

2019.3 성신여자대학교 통계학과 입학

2020.3 정보시스템공학과 복수전공

# Internship

2021.9 ~2021,12

양주시청 기획재정부

정보통신과 빅데이터 인턴

# Certificate

2020.8 컴퓨터 활용능력 1급

2020.8 사회조사분석사 2급

2021.4 데이터 준전문가(ADSP)

# **Project**

2021.2 신용카드 거래 데이터 분석

2021.5 날씨 빅데이터 콘텐트(기상청)

2021,8 서울 열선 도선 우선 설치지 분석

2021.9 어린이 교통사고 취약지 분석

2021.10 폐렴구균 미접종자 데이터 분석

2021.11 민원데이터 분석

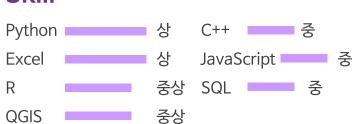
2021.12 지역화폐 및 신용카드 데이터 분석

# Club acitivty

2020.3~7 Smarter 20기 미디어홍보팀

2021.2~ Daja Vu 빅데이터 분석 동아리

# Skill



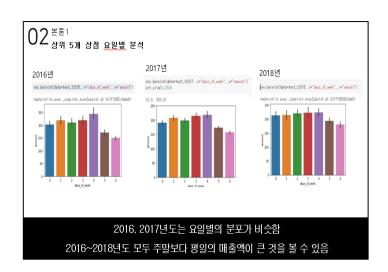
# 신용카드 거래 데이터 분석

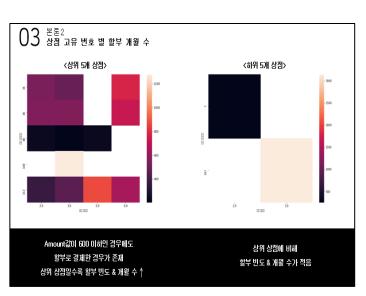
소상공인 신용카드 거래 데이터를 사용하여 분석 패키지를 사용하여 인사이트 발굴 및 시각화

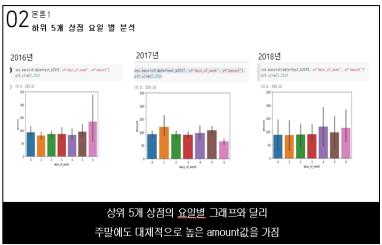
주체기관DACON, 뱅크샐러드분석기간2021.1.18 ~ 2021.2.19기여도<br/>기획80<br/>70

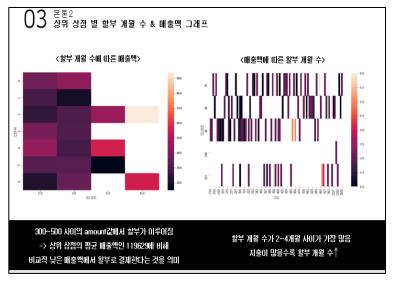
## 분석결과

총 매출을 기준으로 상위.하위 군집을 만듬.매출액에 영향을 미치는 독립변수(할부 개월, 날짜)를 파악하여 군집간의 특성을 비교함. 상위 그룹은 주말보다 평일의 매출액이 높고, 하위 상점은 주말에 높은 매출액이 나타난 걸 파악함. 또한 할부 개월수에 따라 매출액이 달라지는 것을 파악함. 추가적으로 시계열 그래프로 매출액의 변화를 표현함.

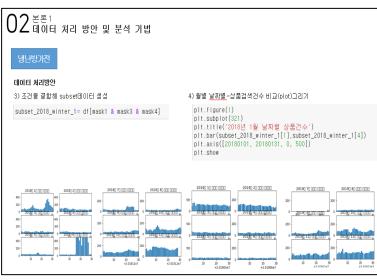


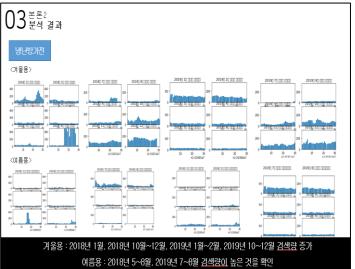


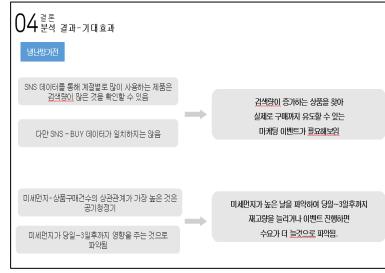












# 날씨 빅데이터 콘테스트

날씨에 따른 소비패턴, 날씨에 민감한 상품군 분석, 날씨에 따른 수요 예측 등 분석을 통해 결과 활용 방안 제시



## 분석결과

대분류가 '냉난방가전 ' 부분을 따로 맡아 분석하였다. 겨울용, 여름용, 사계절용으로 제품을 재분류 한 후 SNS, BUY 데이터를 각각 가설검정(재분류한 계절에 수요가 가장 높은 것) 한 후 월별로 Plot을 그려 수요를 파악함. 여름용, 겨울용 제품은 실제 검색양과 수요량이 비슷한 패턴으로 가는 것을 확인함. 사계절용은 특정 제품이 특정 월에 검색양이 높은 것을 파악. 미세먼지와 상관관계가 가장 높은 '공기청정기'를 뽑아 상관분석을 통해 영향을 끼친다는 것을 파악한 후 이를 통한 마케팅 방안 제시함.

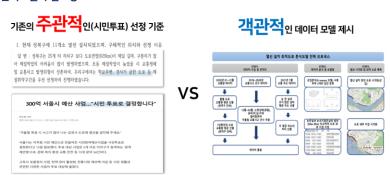
# 서울시 열선 도로 우선 입지 선정

서울시 겨울철 도로 사고 위험도 분석을 통해 열선 도로 우선 입지 선정/공급을 통해 겨울철 교통사고 문제 해결 및 예방

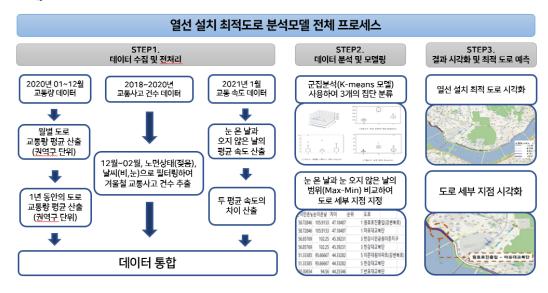
분석 배경 및 목적



## 분석모델의 필요성



## 분석 프로세스



## 분석 배경 및 목적

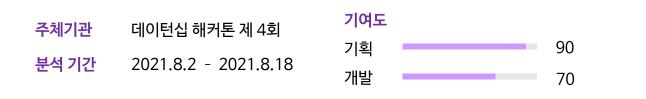
올해 초 **서울시는 폭설 대응에 실패**하여 극심한 교통체증과 교통사고 발생한 반면,

열선 도로 최다 설치구인 성북구는 폭설에도 차량 소통이 원활함

기존에는 입지 선정을 주민들의 민원에 기반하여 열선 도로를 설치했지만

데이터 분석을 통해 서울시 주요 도로 중 우선 설치 도로를 선정.

열선 도로 신규 설치 지점 선정 근거 자료로 활용될 뿐만 아니라 스마트 도로 관리 통합 시스템에 기여 가능함.



# 서울시 열선 도로 우선 입지 선정

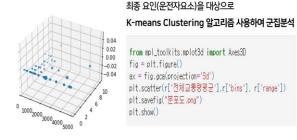
서울시 겨울철 도로 사고 위험도 분석을 통해 열선 도로 우선 입지 선정/공급을 통해 겨울철 교통사고 문제 해결 및 예방

분석 내용 - 군집분석

군집분석 – 변수 선택

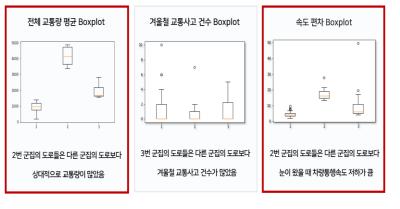


## 클러스터링

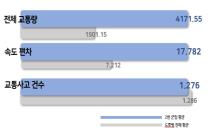


환경 요인(상습결빙구간, 급경사지, 커브기)를 제외한 후 운전자요인(교통량, 사고 건수, 차량 통행 속도)를 선택함. 이를 대상으로 K-means 클러스터링을 알고리즘을 사용하여 군집 분석 실시함. 이후 Box-plot을 통해 변수별 분포를 확인하여 최종 군집 선정

## ₹ 군집분석 - Boxplot

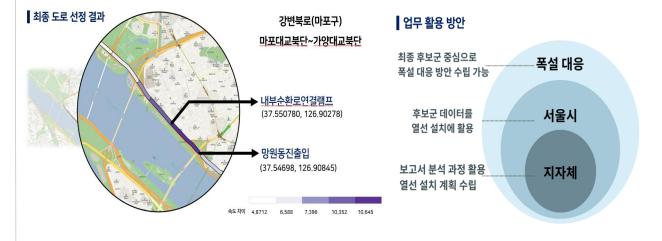


2번 군집은 전체 76개 도로에 비해 교통량이 약 3배 이상 많고 눈 온 날과 안 온 날의 속도 편차도 2배 이상 큰 것을 알 수 있다.



각 도로 구간별로 구분하여 눈이 오지 않은 차량 통행 속도 범위에서 눈이 오는날 차량속도 범위를 제외하여 도로 구간별 차량 통행 속도 범위(MAX-MIN)이 가장 큰 부분을 설치 지점으로 선정함.

## 최종 분석 결과



# 어린이 교통안전 취약지 분석

어린이 교통사고 데이터를 기반으로 머신러닝 회귀분석을 통해 어린이 교통안전 취약지수 수립 및 안전시설물 우선 설치지(보수지) 도출

## 분석 배경 및 목적

# <2020년, 민식이법 시행> (민식이법'이란? 이런이보고구에 대 취소 단속에서 함의 설치와 사망이나 생세수고 가세수에 대한 가중세병 당근 앱이네요.

2019년 김민식 군 사망 이후 민식이법 시행 및 교통사고 강화대책 발표

## <양주시 교통사고 현황>



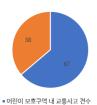
2019-2020년 교통사고 건수 감소 정체

### <경기 지역과 양주시 교통사고 비교 >



어린이 인구 비율이 비슷한 군집끼리 비교시 교통사고가 다소 많음





■ 어린이 보후구역 외 교통사고 건수





●: 교통사고 발생지●: 안전시설물 설치지

양주시에 설치한(설치예정) 안전시설물 대부분은 어린이 보호구역 내에 우선순위로 설치되어 있으나 어린이 교통사고를 줄이기 위해선 <mark>어린이 보호구역 이외에도 교통 안전 취약지에 안전시설물 설치 필요</mark>

## 분석 프로세스

## 어린이 교통안전 취약지역 분석 모델

### 데이터 수집

지자체 데이터 공공데이터

### 공간 데이터 분석

격자 데이터에 독립변수 데이터 매핑

### 머신러닝 회귀 모델

최적 회귀 모델 선택 후 취약지수 산출

### 공간 데이터 분석

취약지수 시각화 현재/잠재적 위험지대 추정

### 분석 내용 - 머신러닝 회귀분석

#### 1. 종속변수 독립변수 선정

종속변수: 어린이 교통사고 건수

Elastic net

Stepwiseline

logistic regressio

독립변수: 종속변수의 유의한 영향을 미칠 것 같은 56개 변수

## 2. 머신러닝 회귀 모델(8가지) 분석





타사고 거소) 표시 심추다



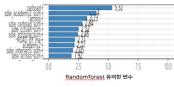
### 3. 최적의 모형 선택

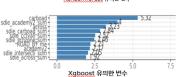






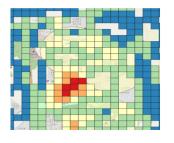
### 4. 변수별 중요도 파악





각각의 모델에서 유의한 변수 뽑은 후 회귀계수 정리

### 5. 현재 교통 안전 취약지 선정



유의한 변수의 회귀계수를 가중치로 두어 결자별 취약지수 값 선정 취약지수가 높은 지역은 빨간색으로 표시



각 모델의 선택된 위험지역 중 겹치는 지역을 최종 현재 교통 안전 취약지로 선정

# 어린이 교통안전 취약지 분석

어린이 교통사고 데이터를 기반으로 머신러닝 회귀분석을 통해 어린이 교통안전 취약지수 수립 및 안전시설물 우선 설치지(보수지) 도출

### 6. 잠재적 교통 안전 취약지 선정

# <유의 변수 도출> Name 설명 academy2 경기도 양주 학원 across 경기도 양주 환단보도 intersection 경기도 양주 교차로 Side\_academy2\_sum 인접격자 경기도 양주 학원 Side across sum 인접격자 경기도 양주 종단보도 Side carboad\_sum 인접격자 경기도 양주 주정차 표지판 지역 Side\_intersection\_sum 인접격자 경기도 양주 교차로 Side\_parkno\_sum 인접격자 경기도 양주 조정차 금지구역 Side schozone\_sum 인접격자 경기도 양주 스물존

각 모델의 유의한 변수중 중복 변수를 필터링하여 최종 유의 변수로 도출함

# 

알맞은 군집개수 선정 후 클러스터링을 통해 유사 지역 분류

### <잠재적 위험지대 확인>



기존 어린이 사고가 일어난 지역을 제외하고 사고가 일어날 지역 선정

## 분석 결과 – 현재 교통 안전 취약지 : 가래비 3길







최근 5년간 교통사고 발생 건수 : 2건

현재 도로 상황 : 교통 통행량이 많지만 횡단보도만 존재.

사각지대로 보행자 및 자전거 탑승 어린이가 차와 충돌하여 사고난 지역

추가(재설치) 안전 시설물: 교통 단속 신호기, 횡단보도 신호등

회귀모델에서 선정된 지역 로드뷰 및 현장답사하여 현재 도로 상황과 필요한 안전시설물 정리하여 해당 부서(교통시설팀, 생활도로팀, 차량안전과)와 협의 후 2022년도 계획안에 반영

## 분석 결과 – 스쿨존 위험지역

머신러닝 세 모델 모두 위험지역이라 선정한 곳과 안전 시설물 미설치 학교를 중심으로 선정하였다.

### 교통 안전 취약지수 상위 초등학교





## 활용 방안

### 어린이 교통사고 발생지 특성 파악 가능

(학원가, 신호등 없는 횡단보도, 교차로, <u>주정차</u> 표지판이나 CCTV 없는 <u>주정차</u> 금지구역 인근 학원가, 인근 스쿨존)

QGIS를 이용한 공간 시각화 결과물을 협업 부서가 활용 가능

### 스쿨존 위험지역 파악 가능

안전 취약지수를 통해 취약지수가 높은 스쿨존을 파악하여 안전시설물 <u>재설치할 스쿨존</u> 선정 가능 예산 효율성 늘 것으로 예상

## b전지 파악 가능 취약지역 규제 단속 및 강화

안전 취약지수를 통해 취약지수가 높은 지역은 교통 안전 규제를 높여 운전자들이 경각심을 가지게 함.

## 현재/잠재적 교통 안전지 파악 가능

빅데이터를 근거로 한 교통 안전 취약지를 파악하여 안전 시설물 우선 설치지 및 재설치지 지역 선정 근거로 활용

주체기관양주시청분석 기간2021.9.6 - 2021.10.15



# 코로나19 중증환자 예방을 위한

# 폐렴구균 미접종자 데이터 분석

폐렴구균 미접종자 취약 연령대 및 취약지를 분석하여 폐렴구균 접종필수라는 인지도 향상을 위해 홍보 방안 및 2022년 접종 계획 반영

## 분석 배경 및 목적

### 폐렴구균 접종확대의 필요성





현재 양주시 만 65세 이상 어르신 폐렴구균 백신 미접종자 수는 10,492명으로 전체 인구 대비 27%이다.



양주시는 고령사회로 진입했으며 2026년 만 65세 인구는 3967명으로 2017년 만 65세 인구 1577명보다 2배 더 많을 것으로 예상된다.

추세선도 향후 노인인구가 계속 증가될 것으로 파악된다.

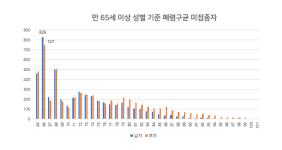
폐렴구균 및 코로자19 동시 감염자는 사망률이 7.8배 증가하며 폐렴환자는 코로나19 중증환자로 분리됨. 현재 양주시 폐렴구균 백신 미접종자수가 27%이며 노인인구가 계속 증가될 것으로 파악됨 따라서 **코로나19 상황에서 폐렴구균 백신접종 중요성 증대** 

주체기관양주시청기획50분석 기간2021.10.20 - 2021.11.5개발90

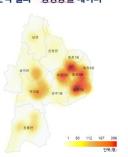
## 분석 결과

### 분석 결과 – 연령별 현황(성별)

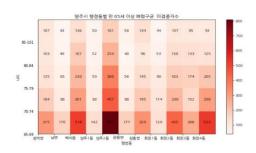




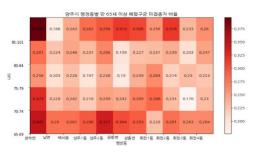
### 분석 결과 – 행정동별 데이터



## 분석 결과 - 히트맵 표현(연령별, 행정동별)



미접종자 수는 <mark>양주2동에 거주하고 있는 66세-70세</mark> 809명으로 제일 많다. 미접종자 수가 가장 많은 <mark>연령은 66-70세,</mark> 가장 많은 <mark>지역은 양주2동</mark>으로 파악된다.



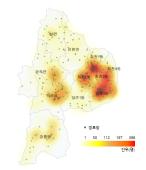
미접종자 비율은 <mark>광적면에 거주하는 85세 이상</mark> 연령층이 0.396으로 제일 높다. 미접종자 비율이 가장 많은 연령은 66-70세, 가장 많은 지역은 <mark>광적면</mark>으로 파악된다.

### 시계열 분석 :결과

연도	예측값	하한값	상한값
2022	1053	911	1199
2023	947	782	1071
2024	1113	838	1164
2025	1110	789	1112
2026	1122	813	1148

결과값으로 향후 5년 만 65세 미접종자 수 예측

### 활용 방안 - 세부 홍보사항(경로당)



QGIS를 통해 양주시 경로당(257개) 반경 300M내에 거주하는 페렴구균 미접종자수 파악

모든 경로당 반경 내 폐렴구균 미접종자 분포



어르신들의 폐렴구균 접종의 필요성에 대한 인지도 향상을 위해

# 민원데이터 분석

민원 빅데이터 분석을 통해 민원 예보 및 조기 대응 등 체계를 정비하고

주기적으로 제기되는 민원의 근본적 해결을 위해 부서별 제도개선에 근거자료로 쓰임.

## 세부 분석 결과



- 민원건수 상위 10개 부서가 전체 민원의 60%를 차지함.
- \*19년도 상반기 81%, 하반기 68%, 20년도 상반기 78%, 하반기 65%에 비해 감소하는 추세를 보임
- 2020년 하반기와 비교하였을 시, 여성보육과와
- 도시발전과가 빠지고 사회복지과와 위생과가 추가됨.
- \*9월 국민지원금과 관련된 사회복지과 민원 증가,
- 코로나 방역수칙 위반과 관련된 꾸준히 민원 증가

- 단기 급증 민원 부서
- 1)사회복지과: 국민지원금(재난지원금)
- 이의신청 건으로 9월 민원 급증
- 2)광역교통시설과: 7호선 104정거장
- 옥정동 출입구 설치요청으로 6-8월민원 증가
- 3) 주택과: 양주 옥정역 대광로제비앙 감리업무 ·
- 시공 부실 철저 감사 및 아파트 공사문의

행 레이블↩	01€	02↩	03↩	04∂	05↩	06↩	07↩	08€	09↩	10↩	총합계실 실
차량관리과리	226€	296€	482€	631€	407€	469	511 <del>4</del>	591+	629€	829€	5071∹∹
복지지원과리	359€	350€	306€	244€	246€	199€	197∉	179€	297∉	300€	2677∹∂
도로과	76€	136€	226∈	217€	165€	166€	197∉	140€	219€	255€	1797⋵∂
주택과↩	169€	107€	196∈	223€	166€	109€	78≓	70€	87∉	57€	1262∹∂
대중교통과↩	66∈	50€	84≓	81€	83€	97∉	101∈	81€	87∉	124	854⋵∂
청소행정과↩	75€	89€	101∈	95€	76⋵	68€	80€	87€	87€	57€	815≓∂
환경관리과↩	48€	35€	37⋵	69€	51∉	113€	135€	129€	93€	90€	800€∂
사회복지과↩	7⋵	9€	7⋵	1€	7≓	107€	0←	9∉	585	59€	791∹≓
광역교통시설과↩	83€	27€	18⋵	20€	16≓	192€	122€	112€	59€	14€	663⋵⋲
위생과↩	53←	48€	43≓	45€	41∈	43€	71≓	65€	23∻	30€	462⋵⋵

## 키워드 분석(LCD분석, 워드 클라우드)

## 민원데이터 텍스트 마이닝 분석 프로세스

### 데이터 전처리

접수요약〉결측값을 제목 칼럼 값으로 대체 접수일자〉'월' 칼럼 생성

### 카테고리 생성 & 불용어 제거

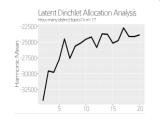
접수요약 칼럼 기준으로 카테고리 생성 <u>관련없는 불용어</u> 제거

### 텍스트 마이닝

LCD 분석을 통해 최적 토픽 도출

### 데이터 시각화

카테고리별 시각화 월별 시각화

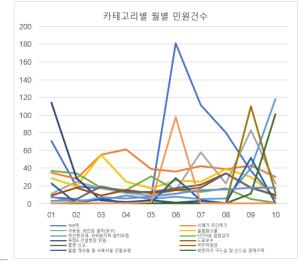






키워드₽

불법주정차



주체기관

양주시청

분석 기간 2021.11.8 - 2021.11.24

기여도

기획 개발 50

70

# 코로나19 전후 양주시 소비 변화 분석

코로나19 전후 양주시 지역 상권 매출 변화(신용카드 및 지역화폐 거래 내역) 및 유동인구 추이를 분석하여 양주시 맞춤형 정책 수립을 위한 기초자료로 활용하고자 함.

세부 분석 결과

# 양주시 신용카드 및 지역화폐 결제금액



- 2020년 총 결제금액은 전년대비 약 2% 증가하였고, 2021년 총 결제 금액은 전년대비 약 8% 증가하였다.
- 총 결제금액의 98.5%이상은 신용카드 결제 금액이다.
- 코로나 발생 전후 전체 결제 금액에 큰 영향은 없었으나, 첫 확진자 발생 및 영업 시간 단축에 따른 매출 감소는 뚜렷했다. (단, 신용카드 데이터는 정책 지원금 여부가 구분되지 않아 지원된 정책 지원금과 분리하여 파악하기 어렵다.)

## 업종 상세 소비

# 업종 상세 변화율

업종코드	2019년(1~12월)	2020년(1~12월)	2021년 (1~10월)
여행/숙박	1,472,171,107,967	1,195,981,513,338	1,092,878,174,697
레져/취미	9,263,522,843,881	8,064,498,715,103	7,591,625,743,403
가구/주유	19,614,042,784,137	18,182,654,042,350	17,505,852,206,592
유통/편의점	56,787,145,593,887	68,016,404,415,552	66,277,084,061,904
교육	9,765,476,091,979	8,638,529,650,681	7,884,516,134,926
자동차/보험	11,529,921,607,351	10,784,751,695,673	8,401,631,787,753
의료	21,895,236,418,385	21,534,624,173,609	19,415,512,750,937
외식/유흥	40,070,528,175,353	36,407,240,232,672	29,720,905,737,701

	2020	2021 (19년대원)
পপ্ত/এখ	-23%	-36%
레저/취미	3%	0%
카구/주유	1%	2%
쇼핑/팬의점	5%	-10%
28	-4%	-11%
자동차/보험	10%	0%
의료	8%	6%
외식/유흥	4%	-9%

- 여행/숙박, 교육 분야의 결제가 감소하였다.
- 2021년 데이터는 10개월의 자료임에도 불구하고 가구/주유, 의료 분야의 결제 금액은 오히려 증가하였다.

# 매출액이 증가한 상위 세부업종 매출액이 감소한 상위 세부업종

	세부업종이름	증가율
2020(19년대비)	가전제품	41%
	음식료품	36%
	직물	28%
	사무/통신기기	28%
	자동차판매	27%
2021(19년대비)	음식료품	78%
	가전제품	54%
	의료기관	39%
	광학제품	29%
	신변잡화	29%

	세부업종이름	감소율
2020(19년대비)	여행업	-61%
	유흥주점	-37%
	광학제품	-18%
	숙박업	-16%
	주방용품	-12%
2021(19년대비)	유흥주점	-81%
	여행업	-74%
	단란주점	-68%
	숙박업	-16%
	회원제형태업소	-8%

주체기관

양주시청

분석 기간 2021.12.6 - 2021.12.31 기여도

기획

50

개발

80