

데이터분석 경진대회

제 9 회 KMU 빅데이터 분석 경진대회 과제 소개

Kim, Yang Sok
MIS @ Keimyung University

목차

- 분석 개요
 - 분석 목적
- 데이터
 - 데이터 셋 1 – 인허가 데이터
 - 데이터 셋 2 – 코로나-19 데이터
- 분석 과제 1, 2, 3
- 분석 결과 제출

분석 개요

- 분석 목적

- 본 분석 경진대회에서는 전국 인허가 데이터 중 숙박업과 관련된 데이터 분석을 목적으로 함.

- 분석 과제

1. 업체 현황 시각화: 지역별로 업체의 현황을 보여줄 수 있는 시각화 작업
2. 폐업 예측: 어떤 업체가 폐업을 했는지 보여주는 작업
3. Covid-19의 영향 분석: Covid-19의 영향으로 인한 폐업의 증가 여부 분석

데이터 셋 1. 인허가 데이터

- 인허가 데이터는 7개 유형의 업종을 포함한 데이터임.

그룹	업소정보
관광숙박업	관광호텔과 같은 숙박시설 및 음식, 운동, 오락 등 적합한 시설 등을 함께 갖추어 제공하는 업소정보
관광펜션업	숙박시설을 운영하며, 자연,문화 체험관광에 적합한 시설을 갖추어 제공하는 업소정보
숙박업	여관, 여인숙, 호텔 등 잠을 자고 머물수 있도록 시설 및 설비 등의 서비스를 제공하는 업소 정보
외국인관광도시민박업	도시 지역의 주민이 거주하고 있는 주택을 이용하여 외국인 관광객에게 한국의 가정 문화를 체험 할 수 있도록 숙식 등의 서비스를 제공하는 업소정보
일반야영장업	야영에 적합한 시설 및 설비 등을 갖추고 야영편의를 제공하는 업소정보
자동차야영장업	자동차를 이용하는 여행자의 야영·취사 및 주차에 적합한 시설을 갖추어 관광객에게 이용을 제공하는 업소정보
한옥체험업	우리나라 고유의 형식으로 지은 한옥에 숙박 체험하기에 적합한 시설을 갖추어 관광객에게 이용을 제공하는 업소정보

데이터 셋 1. 인허가 데이터

- 데이터 셋은 7개의 하위 데이터로 구성되어 있어, 하나의 데이터 셋으로 합쳐서 분석을 수행하여야 함.
- 데이터의 속성 중 “영업상태명”을 라벨(종속데이터)로 사용하여야 함.

붙여넣기

글꼴

맞춤

표시 형식

스타일

셀

편집

11

영업상태명

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	번호	개방서비스	광서비스	광자치단	관리번호	인허가일	허가취소	업상태구분	영업상태	제영업상태	제영업상	폐업일자	휴업시작	휴업종료	재개업일	소재지전	소재지면	재지우편
3	2	관광숙박업	03_11_01_1	3000000	CDFI22600	20180618		03	폐업	03	폐업	20201008						
43	42	관광숙박업	03_11_01_1	3000000	CDFI22600	19670403		02	휴업	02	휴업		20210101	20221231		02-734-68		
44	43	관광숙박업	03_11_01_1	3000000	CDFI22600	19800401		02	휴업	02	휴업		20210210	20230209		02-3676-8		
45	44	관광숙박업	03_11_01_1	3000000	CDFI22600	20140203		03	폐업	03	폐업	20201210						110825
46	45	관광숙박업	03_11_01_1	3000000	CDFI22600	20140214		03	폐업	03	폐업	20200908				02-734-00		110111
47	46	관광숙박업	03_11_01_1	3000000	CDFI22600	20141105		03	폐업	03	폐업	20211229				02-741-38		110450
48	47	관광숙박업	03_11_01_1	3000000	CDFI22600	20151229		03	폐업	03	폐업	20200330				02 22318		
49	48	관광숙박업	03_11_01_1	3000000	CDFI22600	20140122		03	폐업	03	폐업	20210228				070-8146-		110836
50	49	관광숙박업	03_11_01_1	3000000	CDFI22600	19740401		03	폐업	03	폐업	20080104				02-735-90		110130
51	50	관광숙박업	03_11_01_1	3000000	CDFI22600	20030711		03	폐업	03	폐업	20051230				02-2287-6		110846
52	51	관광숙박업	03_11_01_1	3000000	CDFI22600	20031210		03	폐업	03	폐업	20120830				02-763-32		110450
53	52	관광숙박업	03_11_01_1	3000000	CDFI22600	20121127		03	폐업	03	폐업	20161101						110360
54	53	관광숙박업	03_11_01_1	3000000	CDFI22600	20151022		03	폐업	03	폐업	20181031				02-3675-7		
55	54	관광숙박업	03_11_01_1	3000000	CDFI22600	20130512		03	폐업	03	폐업	20201127				070-8958-		110340
56	55	관광숙박업	03_11_01_1	3000000	CDFI22600	20140109		03	폐업	03	폐업	20160215				02-2174-6		110891
153	152	관광숙박업	03_11_01_1	3010000	CDFI22600	19751202		02	휴업	02	휴업		20211220	20221231		2270-3187		100855
154	153	관광숙박업	03_11_01_1	3010000	CDFI22600	19600610		03	폐업	03	폐업	20150731				2268-7111		100260

데이터 셋 2. 코로나 데이터

- 코로나 관련 데이터는 전국 발생 데이터를 기준으로 사용함.
 - 서울시 코로나 관련 데이터는 아래 현황 자료 사용
 - <https://data.seoul.go.kr/dataList/1/literacyView.do>

분석 과제 1 – 시각적 분석

- 업종별, 시도별 인허가 관련 데이터를 시각화하여 현황을 쉽게 이해할 수 있는 시각 분석 자료 작성
- 시계열 현황 자료를 통해 인허가 현황을 파악할 수 있는 시각화 분석 자료 작성
- 코로나 전/후 또는 코로나 발생과 관련하여 현황을 파악할 수 있는 분석 자료 작성

분석 과제 2 – 폐업/휴업 예측

- 데이터 중에서 영업상태명 속성이 영업, 폐업, 휴업 등 3가지 값을 가진 데이터만 활용하여 분류 예측 수행
- 7가지 유형의 업종을 모두 합쳐서 하나의 데이터로 영업/폐업/휴업 분류 예측을 수행
- 데이터를 생성한 후 임의로 데이터를 70:30으로 분할하여 검증 작업 수행함.

분석 과제 3 – 폐업/휴업의 코로나 영향 분석

- 데이터를 월 또는 년 단위로 창업/폐업/휴업 데이터를 생성
- 월 또는 연도별 데이터를 기반으로 2020 및 2021년도 폐업 및 휴업을 예측
- 2020년도 예측과 실제 발생의 차이를 기반으로 코로나의 영향을 분석함.

분석 결과 제출

- 분석 소스 파일(Excel, Python, R, Tableau, Rapidminer)
 - 데이터 전처리 파일
 - 모델링 파일
 - 시각적 분석 파일
- 분석 보고서
 - 분석 프레젠테이션: 파워포인트로 작성, 20 ~ 30 페이지(템플릿 활용)
 - 분석 보고서: Word/HWP 작성, 10 페이지 내외(템플릿 활용)
- 제출 기한: 2020년 2월 4일(금) 오전 10:00
- 제출 방법: yangsokk@gmail.com로 제출하기 바람



QUESTIONS?