

CAPSTONE FOR DATA SCIENCE CERTIFICATION

Analysis of Barcelona's hospitality, hospitality and recreational sector for supply and service industry development

Introduction

This try to be a parcial market study about hospitality sector in Barcelona's city. It's also included the hospitality and recreational business like gyms, in order to support the auxiliar service and utilities industry that supplay to those mentioned main industries.

Target market for the study

For instance, catering companies that work serving to hotels or restaurants, cleaning and desinfection companies, laundries, kitchenware suppliers, fresh food, etc.. all the companies that give services or supply goods for hotels, hostals, restaurants, and/or gyms in Barcelona

Area of study

All the mentioned service's and utilities companies above need to know where fisically their possible objective clients are. Also, they need to create areas dividing the city in order to give the best posible service for their clients. That's why, this studie try to use the known data bases of Barcelona to divide the city in areas and get a possible location for their warehouse or centralization office for each area.

Methodology

It is possible to use K-Mean clustering to divide the city in areas or groups because K-Mean try to group the elements with the less posible distance between elements of the group but getting the highest distance between each group. Furthermore, the K-Mean method provide the centroids, it means, a center point of the group where ideally can be localized an operational centralization like warehouse or office to operate each group. Instead of use an Euclidean space, we can use longitude and latitude as a Cartesian space for K-Means.

Data sources

Barcelona's open data repository

<https://opendata-ajuntament.barcelona.cat/>

In the link above it's possible to find a large source of official information from Barcelon's Council. Thanks to this open data availability, it's why I decided to focus the study in the city. The dataset provides the information in wide range of formats. The acceesibility can be directly through the links in the web or through their different API's. A better explanation of the API management can be found on the link below:

<https://opendata-ajuntament.barcelona.cat/en/desenvolupadors>

Foursquare API

Because in the council dataset, the information about potential recreational business like gyms are mixed with other sectors and the information is not easy to handle, in the study will be used Foursquare to get the closed gyms to the centroids based on their distance. To get it, the Python Foursquare package will be used to query an easy access.

<https://syg1.org/project/foursquare/>

Data Processing

Required libraries

```
python version : 3.8.6
pandas version : 1.1.4
numpy version : 1.19.3
futures version : 4.0.11.0
matplotlib version : 3.3.2
foursquare version : 1.12020.1.30
sklearn version : 0.23.2
geopy version : 2.0.0
```

Data collection

Please, for further information, see the link below to get an explanation of how to get the information from Barcelona's open data repository. For this study it's treating the information of guesthouses, hostels, hotels and restaurants. That information is available in different formats, in this case, the CSV format will be selected. The code will generate a pandas dataframe for each table and print the name of the dataframe and its shape to get an idea about how long the table it's.

<https://opendata-ajuntament.barcelona.cat/en/desenvolupadors>

DataFrame guesthouse shape : (580, 33)

DataFrame hostel shape : (564, 33)

DataFrame hotel shape : (884, 33)

DataFrame Neighborhood_BCN shape : (73, 4)

Format tables and depreciation of unnecessary columns

Accommodations

"guesthouse", "hostel" and "hotel" tables contain the same structure so all of them can be transformed and join in a single dataframe

A "Category" column is created to define the different type of business to keep that information after being joined

In "...NIVELL" columns it's possible to find which category of accommodation it is. For instance, "3ER_NIVELL" column contains the hotel category stars. "2N_NIVELL" column contains if the guesthouse is an apartment, students residency, young hostel or other type. For hostels there isn't any farther information. What it's proposed is to map that information with numbers from 10 to 0. On one hand that information can be categorized, making luxury hotels more interesting than young hostels. On the other hand, it's better for an easier access in further analysis.

Definitions :

- For hotels, stars +10 --> from 0 starts to 5 starts = 5 to 10.
- hostels = 4
 - guesthouse:
 - Apartments = 3
 - student housing and University residences = 2
 - young hostel = 1
 - not defined categories = 0

Now the tables are joined in one dataframe called 'accommodation'

accommodations dataframe

CODI_EQUIPAMENT	EQUIPAMENT	SECCIO	TIPUS_VIA	NOM_CARRER	NUM_CARRER	NUM_CARRER_1	NUM_CARRER_2	CODI_BARRI	NUM_BARRI
0	146175914	Residència Universitària Lleseps	#	PI	Lleseps	12	12.0	31	la Vila de Gràcia
1	146175914	Residència Universitària Lleseps	#	PI	Lleseps	12	12.0	31	la Vila de Gràcia
2	146175914	Residència Universitària Lleseps	#	PI	Lleseps	12	12.0	31	la Vila de Gràcia

3 rows x 35 columns

Unnecessary columns and possible duplicated lines are dropped

Columns before cleaning : Index(['CODI_EQUIPAMENT', 'EQUIPAMENT', 'SECCIO', 'TIPUS_VIA', 'NOM_CARRER', 'NUM_CARRER', 'NUM_CARRER_1', 'NUM_CARRER_2', 'NUM_CARRER_3', 'NUM_CARRER_4', 'NUM_CARRER_5', 'NUM_CARRER_6', 'NUM_CARRER_7', 'NUM_CARRER_8', 'NUM_CARRER_9', 'NUM_CARRER_10', 'NUM_CARRER_11', 'NUM_CARRER_12', 'NUM_CARRER_13', 'NUM_CARRER_14', 'NUM_CARRER_15', 'NUM_CARRER_16', 'NUM_CARRER_17', 'NUM_CARRER_18', 'NUM_CARRER_19', 'NUM_CARRER_20', 'NUM_CARRER_21', 'NUM_CARRER_22', 'NUM_CARRER_23', 'NUM_CARRER_24', 'NUM_CARRER_25', 'NUM_CARRER_26', 'NUM_CARRER_27', 'NUM_CARRER_28', 'NUM_CARRER_29', 'NUM_CARRER_30', 'NUM_CARRER_31', 'NUM_CARRER_32', 'NUM_CARRER_33', 'NUM_CARRER_34', 'NUM_CARRER_35', 'NUM_CARRER_36', 'NUM_CARRER_37', 'NUM_CARRER_38', 'NUM_CARRER_39', 'NUM_CARRER_40', 'NUM_CARRER_41', 'NUM_CARRER_42', 'NUM_CARRER_43', 'NUM_CARRER_44', 'NUM_CARRER_45', 'NUM_CARRER_46', 'NUM_CARRER_47', 'NUM_CARRER_48', 'NUM_CARRER_49', 'NUM_CARRER_50', 'NUM_CARRER_51', 'NUM_CARRER_52', 'NUM_CARRER_53', 'NUM_CARRER_54', 'NUM_CARRER_55', 'NUM_CARRER_56', 'NUM_CARRER_57', 'NUM_CARRER_58', 'NUM_CARRER_59', 'NUM_CARRER_60', 'NUM_CARRER_61', 'NUM_CARRER_62', 'NUM_CARRER_63', 'NUM_CARRER_64', 'NUM_CARRER_65', 'NUM_CARRER_66', 'NUM_CARRER_67', 'NUM_CARRER_68', 'NUM_CARRER_69', 'NUM_CARRER_70', 'NUM_CARRER_71', 'NUM_CARRER_72', 'NUM_CARRER_73', 'NUM_CARRER_74', 'NUM_CARRER_75', 'NUM_CARRER_76', 'NUM_CARRER_77', 'NUM_CARRER_78', 'NUM_CARRER_79', 'NUM_CARRER_80', 'NUM_CARRER_81', 'NUM_CARRER_82', 'NUM_CARRER_83', 'NUM_CARRER_84', 'NUM_CARRER_85', 'NUM_CARRER_86', 'NUM_CARRER_87', 'NUM_CARRER_88', 'NUM_CARRER_89', 'NUM_CARRER_90', 'NUM_CARRER_91', 'NUM_CARRER_92', 'NUM_CARRER_93', 'NUM_CARRER_94', 'NUM_CARRER_95', 'NUM_CARRER_96', 'NUM_CARRER_97', 'NUM_CARRER_98', 'NUM_CARRER_99', 'NUM_CARRER_100', 'NUM_CARRER_101', 'NUM_CARRER_102', 'NUM_CARRER_103', 'NUM_CARRER_104', 'NUM_CARRER_105', 'NUM_CARRER_106', 'NUM_CARRER_107', 'NUM_CARRER_108', 'NUM_CARRER_109', 'NUM_CARRER_110', 'NUM_CARRER_111', 'NUM_CARRER_112', 'NUM_CARRER_113', 'NUM_CARRER_114', 'NUM_CARRER_115', 'NUM_CARRER_116', 'NUM_CARRER_117', 'NUM_CARRER_118', 'NUM_CARRER_119', 'NUM_CARRER_120', 'NUM_CARRER_121', 'NUM_CARRER_122', 'NUM_CARRER_123', 'NUM_CARRER_124', 'NUM_CARRER_125', 'NUM_CARRER_126', 'NUM_CARRER_127', 'NUM_CARRER_128', 'NUM_CARRER_129', 'NUM_CARRER_130', 'NUM_CARRER_131', 'NUM_CARRER_132', 'NUM_CARRER_133', 'NUM_CARRER_134', 'NUM_CARRER_135', 'NUM_CARRER_136', 'NUM_CARRER_137', 'NUM_CARRER_138', 'NUM_CARRER_139', 'NUM_CARRER_140', 'NUM_CARRER_141', 'NUM_CARRER_142', 'NUM_CARRER_143', 'NUM_CARRER_144', 'NUM_CARRER_145', 'NUM_CARRER_146', 'NUM_CARRER_147', 'NUM_CARRER_148', 'NUM_CARRER_149', 'NUM_CARRER_150', 'NUM_CARRER_151', 'NUM_CARRER_152', 'NUM_CARRER_153', 'NUM_CARRER_154', 'NUM_CARRER_155', 'NUM_CARRER_156', 'NUM_CARRER_157', 'NUM_CARRER_158', 'NUM_CARRER_159', 'NUM_CARRER_160', 'NUM_CARRER_161', 'NUM_CARRER_162', 'NUM_CARRER_163', 'NUM_CARRER_164', 'NUM_CARRER_165', 'NUM_CARRER_166', 'NUM_CARRER_167', 'NUM_CARRER_168', 'NUM_CARRER_169', 'NUM_CARRER_170', 'NUM_CARRER_171', 'NUM_CARRER_172', 'NUM_CARRER_173', 'NUM_CARRER_174', 'NUM_CARRER_175', 'NUM_CARRER_176', 'NUM_CARRER_177', 'NUM_CARRER_178', 'NUM_CARRER_179', 'NUM_CARRER_180', 'NUM_CARRER_181', 'NUM_CARRER_182', 'NUM_CARRER_183', 'NUM_CARRER_184', 'NUM_CARRER_185', 'NUM_CARRER_186', 'NUM_CARRER_187', 'NUM_CARRER_188', 'NUM_CARRER_189', 'NUM_CARRER_190', 'NUM_CARRER_191', 'NUM_CARRER_192', 'NUM_CARRER_193', 'NUM_CARRER_194', 'NUM_CARRER_195', 'NUM_CARRER_196', 'NUM_CARRER_197', 'NUM_CARRER_198', 'NUM_CARRER_199', 'NUM_CARRER_200', 'NUM_CARRER_201', 'NUM_CARRER_202', 'NUM_CARRER_203', 'NUM_CARRER_204', 'NUM_CARRER_205', 'NUM_CARRER_206', 'NUM_CARRER_207', 'NUM_CARRER_208', 'NUM_CARRER_209', 'NUM_CARRER_210', 'NUM_CARRER_211', 'NUM_CARRER_212', 'NUM_CARRER_213', 'NUM_CARRER_214', 'NUM_CARRER_215', 'NUM_CARRER_216', 'NUM_CARRER_217', 'NUM_CARRER_218', 'NUM_CARRER_219', 'NUM_CARRER_220', 'NUM_CARRER_221', 'NUM_CARRER_222', 'NUM_CARRER_223', 'NUM_CARRER_224', 'NUM_CARRER_225', 'NUM_CARRER_226', 'NUM_CARRER_227', 'NUM_CARRER_228', 'NUM_CARRER_229', 'NUM_CARRER_230', 'NUM_CARRER_231', 'NUM_CARRER_232', 'NUM_CARRER_233', 'NUM_CARRER_234', 'NUM_CARRER_235', 'NUM_CARRER_236', 'NUM_CARRER_237', 'NUM_CARRER_238', 'NUM_CARRER_239', 'NUM_CARRER_240', 'NUM_CARRER_241', 'NUM_CARRER_242', 'NUM_CARRER_243', 'NUM_CARRER_244', 'NUM_CARRER_245', 'NUM_CARRER_246', 'NUM_CARRER_247', 'NUM_CARRER_248', 'NUM_CARRER_249', 'NUM_CARRER_250', 'NUM_CARRER_251', 'NUM_CARRER_252', 'NUM_CARRER_253', 'NUM_CARRER_254', 'NUM_CARRER_255', 'NUM_CARRER_256', 'NUM_CARRER_257', 'NUM_CARRER_258', 'NUM_CARRER_259', 'NUM_CARRER_260', 'NUM_CARRER_261', 'NUM_CARRER_262', 'NUM_CARRER_263', 'NUM_CARRER_264', 'NUM_CARRER_265', 'NUM_CARRER_266', 'NUM_CARRER_267', 'NUM_CARRER_268', 'NUM_CARRER_269', 'NUM_CARRER_270', 'NUM_CARRER_271', 'NUM_CARRER_272', 'NUM_CARRER_273', 'NUM_CARRER_274', 'NUM_CARRER_275', 'NUM_CARRER_276', 'NUM_CARRER_277', 'NUM_CARRER_278', 'NUM_CARRER_279', 'NUM_CARRER_280', 'NUM_CARRER_281', 'NUM_CARRER_282', 'NUM_CARRER_283', 'NUM_CARRER_284', 'NUM_CARRER_285', 'NUM_CARRER_286', 'NUM_CARRER_287', 'NUM_CARRER_288', 'NUM_CARRER_289', 'NUM_CARRER_290', 'NUM_CARRER_291', 'NUM_CARRER_292', 'NUM_CARRER_293', 'NUM_CARRER_294', 'NUM_CARRER_295', 'NUM_CARRER_296', 'NUM_CARRER_297', 'NUM_CARRER_298', 'NUM_CARRER_299', 'NUM_CARRER_300', 'NUM_CARRER_301', 'NUM_CARRER_302', 'NUM_CARRER_303', 'NUM_CARRER_304', 'NUM_CARRER_305', 'NUM_CARRER_306', 'NUM_CARRER_307', 'NUM_CARRER_308', 'NUM_CARRER_309', 'NUM_CARRER_310', 'NUM_CARRER_311', 'NUM_CARRER_312', 'NUM_CARRER_313', 'NUM_CARRER_314', 'NUM_CARRER_315', 'NUM_CARRER_316', 'NUM_CARRER_317', 'NUM_CARRER_318', 'NUM_CARRER_319', 'NUM_CARRER_320', 'NUM_CARRER_321', 'NUM_CARRER_322', 'NUM_CARRER_323', 'NUM_CARRER_324', 'NUM_CARRER_325', 'NUM_CARRER_326', 'NUM_CARRER_327', 'NUM_CARRER_328', 'NUM_CARRER_329', 'NUM_CARRER_330', 'NUM_CARRER_331', 'NUM_CARRER_332', 'NUM_CARRER_333', 'NUM_CARRER_334', 'NUM_CARRER_335', 'NUM_CARRER_336', 'NUM_CARRER_337', 'NUM_CARRER_338', 'NUM_CARRER_339', 'NUM_CARRER_340', 'NUM_CARRER_341', 'NUM_CARRER_342', 'NUM_CARRER_343', 'NUM_CARRER_344', 'NUM_CARRER_345', 'NUM_CARRER_346', 'NUM_CARRER_347', 'NUM_CARRER_348', 'NUM_CARRER_349', 'NUM_CARRER_350', 'NUM_CARRER_351', 'NUM_CARRER_352', 'NUM_CARRER_353', 'NUM_CARRER_354', 'NUM_CARRER_355', 'NUM_CARRER_356', 'NUM_CARRER_357', 'NUM_CARRER_358', 'NUM_CARRER_359', 'NUM_CARRER_360', 'NUM_CARRER_361', 'NUM_CARRER_362', 'NUM_CARRER_363', 'NUM_CARRER_364', 'NUM_CARRER_365', 'NUM_CARRER_366', 'NUM_CARRER_367', 'NUM_CARRER_368', 'NUM_CARRER_369', 'NUM_CARRER_370', 'NUM_CARRER_371', 'NUM_CARRER_372', 'NUM_CARRER_373', 'NUM_CARRER_374', 'NUM_CARRER_375', 'NUM_CARRER_376', 'NUM_CARRER_377', 'NUM_CARRER_378', 'NUM_CARRER_379', 'NUM_CARRER_380', 'NUM_CARRER_381', 'NUM_CARRER_382', 'NUM_CARRER_383', 'NUM_CARRER_384', 'NUM_CARRER_385', 'NUM_CARRER_386', 'NUM_CARRER_387', 'NUM_CARRER_388', 'NUM_CARRER_389', 'NUM_CARRER_390', 'NUM_CARRER_391', 'NUM_CARRER_392', 'NUM_CARRER_393', 'NUM_CARRER_394', 'NUM_CARRER_395', 'NUM_CARRER_396', 'NUM_CARRER_397', 'NUM_CARRER_398', 'NUM_CARRER_399', 'NUM_CARRER_400', 'NUM_CARRER_401', 'NUM_CARRER_402', 'NUM_CARRER_403', 'NUM_CARRER_404', 'NUM_CARRER_405', 'NUM_CARRER_406', 'NUM_CARRER_407', 'NUM_CARRER_408', 'NUM_CARRER_409', 'NUM_CARRER_410', 'NUM_CARRER_411', 'NUM_CARRER_412', 'NUM_CARRER_413', 'NUM_CARRER_414', 'NUM_CARRER_415', 'NUM_CARRER_416', 'NUM_CARRER_417', 'NUM_CARRER_418', 'NUM_CARRER_419', 'NUM_CARRER_420', 'NUM_CARRER_421', 'NUM_CARRER_422', 'NUM_CARRER_423', 'NUM_CARRER_424', 'NUM_CARRER_425', 'NUM_CARRER_426', 'NUM_CARRER_427', 'NUM_CARRER_428', 'NUM_CARRER_429', 'NUM_CARRER_430', 'NUM_CARRER_431', 'NUM_CARRER_432', 'NUM_CARRER_433', 'NUM_CARRER_434', 'NUM_CARRER_435', 'NUM_CARRER_436', 'NUM_CARRER_437', 'NUM_CARRER_438', 'NUM_CARRER_439', 'NUM_CARRER_440', 'NUM_CARRER_441', 'NUM_CARRER_442', 'NUM_CARRER_443', 'NUM_CARRER_444', 'NUM_CARRER_445', 'NUM_CARRER_446', 'NUM_CARRER_447', 'NUM_CARRER_448', 'NUM_CARRER_449', 'NUM_CARRER_450', 'NUM_CARRER_451', 'NUM_CARRER_452', 'NUM_CARRER_453', 'NUM_CARRER_454', 'NUM_CARRER_455', 'NUM_CARRER_456', 'NUM_CARRER_457', 'NUM_CARRER_458', 'NUM_CARRER_459', 'NUM_CARRER_460', 'NUM_CARRER_461', 'NUM_CARRER_462', 'NUM_CARRER_463', 'NUM_CARRER_464', 'NUM_CARRER_465', 'NUM_CARRER_466', 'NUM_CARRER_467', 'NUM_CARRER_468', 'NUM_CARRER_469', 'NUM_CARRER_470', 'NUM_CARRER_471', 'NUM_CARRER_472', 'NUM_CARRER_473', 'NUM_CARRER_474', 'NUM_CARRER_475', 'NUM_CARRER_476', 'NUM_CARRER_477', 'NUM_CARRER_478', 'NUM_CARRER_479', 'NUM_CARRER_480', 'NUM_CARRER_481', 'NUM_CARRER_482', 'NUM_CARRER_483', 'NUM_CARRER_484', 'NUM_CARRER_485