COURSERA CAPSTONE PROJECT
IBM APPLIED DATA SCIENCE CAPSTONE

OPENING A TRULY ITALIAN RESTAURANT IN MILAN

GIUSEPPE DE NITTIS

coursera

IBM

BUSINESS PROBLEM

❖ FOOD IS ONE OF THE MOST IMPORTANT RESOURCES FOR MANKIND FOR ITALIANS, IT IS EVEN MORE IMPORTANT

❖ DECIDING TO OPEN A RESTAURANT IN ITALY CAN RESULT IN A VERY COMPLEX DECISION

❖ THE OBJECTIVE OF THIS CAPSTONE PROJECT IS TO DECIDE THE BEST LOCATION TO OPEN A NEW RESTAURANT IN THE CITY OF MILAN (ITALY)

❖ AUDIENCE: INVESTORS WILLING TO OPEN A NEW CLASSY PLACE, A UNIQUE SPOT IN THE CITY, WITH HIGH-LEVEL CHEFS AND SOMMELIERS IN ONE OF THE MOST IMPORTANT ITALIAN CITIES

❖ BUSINESS QUESTION: WHERE WOULD YOU OPEN A NEW TRULY ITALIAN RESTAURANT IN THE CITY OF MILAN?

DATA

DATA REQUIRED

❖ List of districts in Milan

❖ Geographical coordinates (namely, latitudes and longitudes) of Milan districts

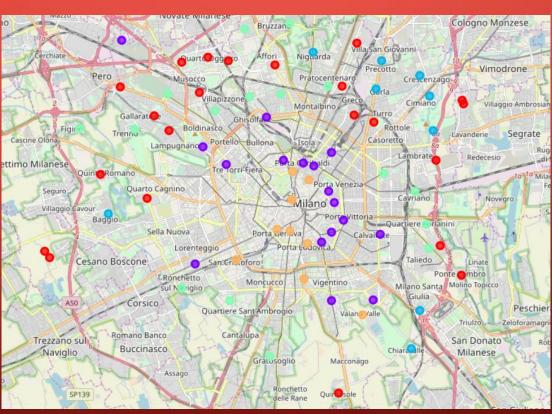❖ Venue data, specifically related to Italian restaurants

DATA SOURCES

❖ Wikipedia page for Milan districts: https://en.wikipedia.org/wiki/Category:Districts_of_Milan

❖ Geocoder package for latitude and longitude coordinates

❖ Foursquare APIs to retrieve the venue data

METHODOLOGY

❖ WEB-SCRAPE WIKIPEDIA PAGE TO GET THE LIST OF MILAN DISTRICTS
❖ USE GEOCODER TO GET LATITUDE AND LONGITUDE COORDINATES OF THE DISTRICTS
❖ USE FOURSQUARE APIS TO GET VENUE DATA
❖ GROUP DATA BY DISTRICT AND TAKE THE MEAN OF THE FREQUENCY OF OCCURRENCE OF EACH VENUE CATEGORY
❖ FILTER VENUE CATEGORY BY *ITALIAN RESTAURANTS*
❖ PERFORM CLUSTERING ON THE DATA BY USING K-MEANS CLUSTERING
❖ VISUALIZE THE CLUSTERS IN A MAP USING FOLIUM

RESULTS

❖ CLUSTER 0 IS SHOWN IN RED
❖ CLUSTER 1 IS SHOWN IN BLUE
❖ CLUSTER 2 IS SHOWN IN DARK GREEN
❖ CLUSTER 3 IS SHOWN IN LIGHT GREEN
❖ CLUSTER 4 IS SHOWN IN ORANGE

DISCUSSION - 1

❖ CLUSTERS 0, 2 AND 3 CONTAIN THE HIGHEST PERCENTAGE OF RESTAURANTS. THIS IS DUE TO THE FACT THAT MOST OF THESE DISTRICTS ARE BIG IN TERMS OF AREA AND OUTSIDE FROM THE VERY CENTER OF THE CITY.

❖ CLUSTERS 1 AND 4, DESPITE BEING IN THE CENTER, PRESENT VALUES THAT ARE SIGNIFICANTLY SMALLER IF COMPARED TO THE ONES OF THE PREVIOUS CLUSTERS. THUS, THE PRESENCE IF ITALIAN RESTAURANTS IS NOT HIGH SINCE THERE ARE A LOT OF DIFFERENT TYPES OF RESTAURANTS

❖ IN THE OUTER DISTRICTS, THE PRESENCE OF ITALIAN RESTAURANTS IS HIGHER SINCE THEY ARE MORE RESIDENTIAL AREAS, WHERE COMMUTERS AND MOSTLY ITALIANS LIVE, SUCH RESTAURANTS CAN BE FOUND EASIER

DISCUSSION - II

❖ THE BEST THING TO DO IS TO OPEN A RESTAURANT IN THE CENTRAL AREA OF MILAN

❖ LOOKING AT THE MAP, NOTICE THAT IN THE EASTERN AREA THERE ARE NO POINTS. IN FACT, THIS AREA HAVE A VERY IMPORTANT MUSEUM AND A UNIVERSITY, THUS THERE ARE MORE BARS. MOREOVER, IT IS ALSO A CHIC DISTRICT, SO OPENING A FANCY AND CLASSY RESTAURANT HERE WOULD BE A GOOD CHOICE

CONCLUSIONS AND FUTURE

❖ THIS PROJECT HAS BEEN DEVELOPED TO ANSWER THE FOLLOWING QUESTION: WHERE WOULD YOU OPEN A NEW TRULY ITALIAN RESTAURANT IN THE CITY OF MILAN?

❖ THE RESULTS I FOUND ARE INTERESTING: OPENING AN ITALIAN RESTAURANT IN THE CENTER OF ONE OF THE BIGGEST ITALIAN CITIES MAY SEEM COUNTERINTUITIVE, BUT IT WOULD APPEAR AS THE BEST CHOICE

❖ FUTURE DIRECTIONS: HERE, I JUST USED FOURSQUARE DATA TO DETERMINE THE BEST LOCATION, BUT MORE FACTORS SHOULD BE INCLUDED, E.G., INFORMATION ON THE PEOPLE LIVING IN THE DIFFERENT DISTRICTS

THANKS FOR YOUR ATTENTION