1. Identifying Fast Food Restaurants in NYC Dataset: https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j? fbclid=lwAR3CJTb7mfSK1oNklcONEyH6fdprYXvrMS4a6S71ri8y98D4NvBoOzpKY24 In [135... import pandas as pd import os import re In [136... os.getcwd() 'C:\\Users\\linel\\Desktop\\Spec_Graphics\\d7_supermarkets_workspace' Out[136... Read in and view data Data dictionary available at https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j? fbclid=lwAR3CJTb7mfSK1oNklcONEyH6fdprYXvrMS4a6S71ri8y98D4NvBoOzpKY24 Explanation of columns: CAMIS: This is an unique identifier for the entity (restaurant); 10-digit integer, static per restaurant permit DBA: "doing business as" (storefront name) BORO: borough- 1 = MANHATTAN • 2 = BRONX • 3 = BROOKLYN • 4 = QUEENS • 5 = STATEN ISLAND BUILDING: building number STREET: street name ZIPCODE: zipcode PHONE: Phone number provided by restaurant owner/manager CUISINE DESCRIPTION: Optional field provided by provided by restaurant owner/manager INSPECTION DATE: NOTE: Inspection dates of 1/1/1900 mean an establishment has not yet had an inspection COUNCIL DISTRICT: Important, we will use this to locate only those in D7 GDrive link to DOHMH_New_York_City_Restaurant_Inspection_Results.csv: https://drive.google.com/file/d/1kHRsENDymawV-XkccDhK4fmFkw68dFSk/view?usp=sharing In [137... restaurants df = pd.read csv("./DOHMH New York City Restaurant Inspection Results.csv") #restaurants df restaurants df.shape (381502, 26)Out[137... **Select only District 7 Restaurants** Leaves us with 6220 of the original 380k NYC restaurants In [138... d7 restaurants df = restaurants df.loc[restaurants df["Council District"] == 7] #d7 restaurants df #d7 restaurants df.shape What kind of DBA names do we have? Cuisine Descriptions? (will help us filter out fast food) In [139... d7 restaurants df["DBA"] JOHN'S FRIED CHICKEN Out[139... BIRCH COFFEE 175 AWASH ETHIOPIAN RESTAURANT 224 THE FAMOUS JIMBO'S HAMBURGER PALACE HUNAN CHEN'S KITCHEN 381295 381409 DEMITASSE COFFEE & TEA 381435 GEISHA ASIAN CUISINE & LOUNGE 381436 GEISHA ASIAN CUISINE & LOUNGE 381442 STROKOS GOURMET DELI Name: DBA, Length: 6220, dtype: object In [140... d7 restaurants df["CUISINE DESCRIPTION"] Chicken Out[140... Coffee/Tea 175 Ethiopian 224 Hamburgers Indian Chinese 381295 381409 Coffee/Tea Asian/Asian Fusion 381435 381436 Asian/Asian Fusion 381442 Sandwiches Name: CUISINE DESCRIPTION, Length: 6220, dtype: object Top 50 US fast-food restaurant chains by 2020 sales: https://www.qsrmagazine.com/content/qsr50-2021-top-50-chart Cross-referenced with https://www.restaurantbusinessonline.com/top-500-chains?year=2021&page=1#data-table In [141... top 50 restaurant chains = ["McDonald's", "Starbucks", "Chick-fil-A", "Taco Bell", "Wendy's", "Burger King", "I "Subway", "Domino's", "Chipotle Mexican Grill", "Sonic Drive-In", "Panera Bread", "Pizza Hut", "KFC", "Popeyes Louisiana Kitchen", "Arby's", "Little Caesars", "Dairy Queer "Panda Express", "Jack in the Box", "Papa John's", "Whataburger", "Wingstop", "Zaxby's", "Hardee's", "Culver's", "Jimmy John's", "Raising Cane's Chicken Fingers", "Five Guys", "Jersey Mike's Subs", "Carl's Jr.", "Bojangles", "In-N-Out Burger", "El Pollo Loco", "Del Taco", "Checkers Drive-In Restaurants", "Firehouse Subs", "Krispy Kreme", "Papa Murphy's Pizza", "Steak 'n Shake", "Qdoba", "Marco's Pizza", "Church's Chicken", "Tim Hortons", "Tropical Smoothie Cafe", "Freddy's Frozen Custard and Steakburgers", "Mcalister's Deli", "Baskin Robbins", "White Castle", "Moe's"] FAST FOOD KEYWORDS = list() for chain in top 50 restaurant chains: FAST FOOD KEYWORDS.append(chain.lower()) #FAST FOOD KEYWORDS len(FAST FOOD KEYWORDS) Out[141... In [142... def isFastFood(business name): if type(business name) != str: return False # in case of empty entries business name = business name.lower() businesses list = business name.split(",") # some locations have multiple names listed for keyword in FAST FOOD KEYWORDS: for business in businesses list: if business in keyword: return True return False In [143... ## Example: ## isFastFood("MCDONALD'S") Determine if each D7 restaurant is a fast-food place We'll mostly focus on filtering out big national fast-food chains In [144... is fast food list = list() dba_list = d7_restaurants_df["DBA"].tolist() #dba list #type(dba list[0]) In [145... for dba in dba list: is fast food list.append(isFastFood(dba)) In [146... #is_fast_food_list In [147... d7_restaurants_fast_df = d7_restaurants_df.assign(fast_food = is_fast_food_list) #d7_restaurants_fast_df In [148... ## Keep only fast food restaurants and remove unnecessary columns d7_restaurants_fast_df = d7_restaurants_fast_df.loc[d7_restaurants_fast_df["fast_food"] == True] d7_restaurants_fast_df = d7_restaurants_fast_df.set_index("CAMIS") d7_restaurants_fast_df = d7_restaurants_fast_df.drop_duplicates() d7_restaurants_fast_df = d7_restaurants_fast_df[~d7_restaurants_fast_df.index.duplicated(keep='first')] d7_restaurants_fast_df = d7_restaurants_fast_df.drop(columns=['BORO','PHONE','CUISINE DESCRIPTION', 'INSPECTION DATE', 'VIOLATION CODE', 'VIOLATION DESCRIPTION', 'CRITICAL FLAG', 'SCORE', 'GRADE', 'GRADE DATE', 'INSPECTION TYPE', 'RECORD DATE', 'Community Board', 'Census Tract', 'BIN', 'BBL', 'NTA', 'ACTION', 'Council District', 'fast food']) In [149.. d7 restaurants fast df Out[149... **DBA BUILDING** STREET ZIPCODE Latitude Longitude **CAMIS** 40605511 DOMINO'S 409 WEST 125 STREET 10027.0 40.811588 -73.954511 41241639 MCDONALD'S 10031.0 40.826699 3549 **BROADWAY** -73.950259 50049533 **DUNKIN** 2568 **BROADWAY** 40.795012 -73.971501 40.806956 40736137 **STARBUCKS BROADWAY** 10025.0 2929 -73.964672 41406255 TACO BELL, KFC **BROADWAY** 40.829692 -73.948081 3645 CHIPOTLE MEXICAN GRILL 50104804 3781 **BROADWAY** 10032.0 40.834151 -73.944822 50048555 **SUBWAY BROADWAY** 40.826913 -73.950103 3559 10031.0 50043183 CHIPOTLE MEXICAN GRILL 805 **COLUMBUS AVENUE** 10025.0 40.794981 -73.965957 41704745 **SUBWAY** 3419 **BROADWAY** 40.822547 -73.953293 10031.0 50008136 WENDY'S 3939 **BROADWAY** 10032.0 40.838876 -73.941366 50044842 **DUNKIN, BASKIN ROBBINS** SAINT NICHOLAS AVENUE 40.830998 -73.941319 10032.0 50018303 WEST 104 STREET 40.799282 **DUNKIN** 206 10025.0 -73.967053 50011485 **SUBWAY** 3795 **BROADWAY** 10032.0 40.834491 -73.944572 50071242 **SUBWAY** 971 AMSTERDAM AVENUE 10025.0 40.801439 -73.964960 41293203 **STARBUCKS BROADWAY** 40.799024 2690 10025.0 -73.968581 41704598 **DUNKIN BROADWAY** 40.821082 3369 10031.0 -73.954360 41716092 DOMINO'S 3624 **BROADWAY** 40.829154 10031.0 -73.948443 41271036 CHIPOTLE MEXICAN GRILL 2843 **BROADWAY** 10025.0 40.804294 -73.966617 50003735 **DUNKIN BROADWAY** 40.835693 -73.943696 3833 10032.0 50041218 LITTLE CAESARS **BROADWAY** 40.835665 3830 10032.0 -73.943689 50046844 DOMINO'S AMSTERDAM AVENUE 965 10025.0 40.801293 -73.965065 41679966 **SUBWAY** 3920 10032.0 40.838530 -73.941594 **BROADWAY** 50099039 **DUNKIN** 1235 AMSTERDAM AVENUE -73.958862 50043586 **DUNKIN** 1995 AMSTERDAM AVENUE 10032.0 40.833913 -73.941248 50058953 DOMINO'S 1988 AMSTERDAM AVENUE 40.833776 -73.941374 50005098 **DUNKIN** AMSTERDAM AVENUE 1342 10027.0 40.813644 -73.956078 41611084 MCDONALD'S 3809 **BROADWAY** 40.834969 -73.944221 10032.0 50045121 **SUBWAY** 578 WEST 125 STREET 10027.0 40.815295 -73.957969 41306786 **DUNKIN** 3600 **BROADWAY** 40.828433 -73.948971 10031.0 41211622 MCDONALD'S 2726 **BROADWAY** 10025.0 40.800086 -73.968089 Save output to csv In [150.. d7 restaurants fast df.to csv("fast food in d7.csv") 2. Identifying and Categorizing Convenience Stores v. Grocery Stores in NYC Get dataframe of all food retailers in D7 (narrowed down from https://data.ny.gov/Economic-Development/Retail-Food-Stores-Map/p2dnxhaw using R to superimpose D7 geographical boundaries) Raw retail food stores data (all NY): https://drive.google.com/file/d/1HQaKZQHDrmCigJS6Pi0h-PrpFTxHytTC/view?usp=sharing food_retailers_in_d7.csv: https://drive.google.com/file/d/1S_j6ol0iWeLjehv3cvX7tJ7P6EtZaAdk/view?usp=sharing In [151... food_retailers_d7_df = pd.read_csv("./food_retailers_in_d7.csv") # food retailers d7 df # food retailers_d7_df.shape Get dataframe of all grocery/fresh produce stores in D7 (manually selected by examining business names) grocery_stores_in_d7.csv: https://drive.google.com/file/d/1hhFLg9ns3nbxFPqNPyqer70sXCFe408_/view?usp=sharing In [152.. grocery_stores_d7_df = pd.read_csv("./grocery_stores_in_d7.csv") grocery_stores_d7_df.rename(columns={}) grocery_stores_d7_df Out[152... Operation.Type Street.Name License.Number **Establishment.Type Entity.Name** DBA.Name Street.Number City State New 3550 FOOD **KEY FOOD** NEW 622503 0 3550 **BROADWAY** NY JAC Store York **CORP** SUPERMARKET **MANHATTEN MANHATTEN AVENUE** New WEST 106TH NEW 1 739952 Store JAC **AVE** 51 NY **COMMUNITY** STREET YORK York **COMMUNITY** CAFE I New **ABC MART** NEW 2 725281 Α **ABC MART** 3505 **BROADWAY** NY Store York NC YORK **DDY MINI DDY MINI AMSTERDAM** NEW New MARKET & 3 629213 Store **JAC** MARKET & 955 NY **YORK** York **AVE DELI CORP** DELI HB DELI HB DELI NEW New 4 742768 Store **JAC GROCERY** 3787A **BROADWAY** NY **GROCERY** YORK York **CORP DYLANS DELICIOUS DYLANS** RIVERSIDE NEW New 731926 106 DELI **DELICIOUS** 900 Store JAC NY DRIVE YORK York GROCERY **DELI GROCERY** COR **MILANO MILANO** New NEW 107 2892 624394 **JAC BROADWAY** NY Store York MARKET INC **MARKET YORK** 91 **HAMILTON** CATHERINE 91 CATHERINE New NEW 627088 108 **JAC** 91 Store NY **GROCERY** York **GROCERY** PLYORK **CORP** 3011 New NEW 621015 109 Store JAC **GROCERY** C TOWN 3320 **BROADWAY** NY **YORK** York INC KING MEAT New NEW 110 708796 JAC **SHOP FAIR** 3871 **BROADWAY** Store NY York **CORP YORK** 111 rows × 22 columns Set "grocery" column in food_retailers_d7_df to "true"/"false" based on whether it appears in our manual selection of grocery stores listed in grocery_stores_d7_df. In [153... # Get list of all grocery stores grocery_stores_list = grocery_stores_d7_df["License.Number"].to_list() #grocery_stores list In [154... # Create column indicating if each store is a grocery store; all "A" establishments are NON-grocery stores food_retailers_list = food_retailers_d7_df["License.Number"].to list() establishment_type_list = food_retailers_d7_df["Establishment.Type"].to_list() is_grocery = list() for i in range(len(food_retailers_list)): if (food_retailers_list[i] in grocery_stores_list) and (establishment_type_list[i] != "A"): is_grocery.append(True) else: is grocery.append(False) #is_grocery In [155... # Add new column to dataframe food retailers d7_df["grocery"] = is_grocery In [156.. food_retailers_d7_df License.Number Operation.Type Street.Name Out[156... **Establishment.Type Entity.Name DBA.Name Street.Number** Address.Line.2 A **DOLLAR DOLLAR** New **AMSTERDAM** 0 627413 Α VARIETY **VARIETY** 762 Store NaN **AVE** York STORE INC **STORE GALAXY** New GALAXY 1 **BROADWAY** 717156 JAC **GOURMET** Store 3778 NaN **GOURMET** York INC 3550 FOOD **KEY FOOD** New 2 622503 Store **JAC** 3550 **BROADWAY** NaN CORP SUPERMARKET York **FINEST FINEST AMSTERDAM** New 3 JAC **COLUMBIA COLUMBIA** 945 739966 Store NaN York AVE **DELI INC** DELI **SANTIAGO SANTIAGO** New **BROADWAY** 4 628824 **JAC DELI FOOD** 3415 Store NaN York **DELI FOOD CORP NEW DELI 1** New 227 741584 JAC **NEW DELI 1** 3660 **BROADWAY** Store NaN York **CORP UPPER UPPER AMSTERDAM** New 228 MANHATTAN 748699 **MANHATTAN** 1728 NaN Store **AVE** York PHARMA INC PHARMACY DUANE **DUANE READE** New 229 **READE ETAL** 626558 2864 **BROADWAY** NaN Store York 14225 **PTRS JOES JOES COLUMBUS** New 230 920 721538 Store JAC **GOURMET GOURMET** NaN **AVENUE** York **DELI INC** DELI 1967 99 CENTS & 99 CENTS & New 231 627132 Α 1967 AMSTERDAM Store NaN York DISCOUNT DISCOUNT INC 232 rows × 18 columns Note: All non-grocery stores ("grocery" = False) are pharmacies, tobacco/liquor stores, delis, or other convenience stores that do not sell a wide enough range of fresh produce to be considered grocery stores. In [157... # See the grocery stores food_retailers_d7_df.loc[food_retailers_d7_df["grocery"] == True] Out[157... License.Number Operation.Type Establishment.Type **Entity.Name** DBA.Name Street.Number Street.Name Address.Line.2 A 3550 FOOD **KEY FOOD** New 2 622503 3550 **BROADWAY** Store JAC NaN York CORP SUPERMARKET **MANHATTEN MANHATTEN AVENUE** WEST 106TH New 5 739952 Store JAC **AVE** 51 NaN COMMUNITY **STREET** York **COMMUNITY** CAFE I **DDY MINI DDY MINI AMSTERDAM** New 7 629213 JAC MARKET & MARKET & 955 NaN Store York **DELI CORP** DELI HB DELI HB DELI New 10 742768 JAC **GROCERY** 3787A **BROADWAY** Store NaN York **GROCERY CORP** NY HARVEST **EVERFRESH** New 12 740014 JAC **TRADING** 2840 **BROADWAY** NaN Store York **SUSHI** INC **DYLANS DELICIOUS DYLANS RIVERSIDE** New 215 731926 **JAC DELICIOUS** 900 DELI NaN Store York **DRIVE GROCERY** DELI G COR **MILANO MILANO** New 216 2892 **BROADWAY** 624394 Store NaN MARKET INC **MARKET** York 91 **CATHERINE** 91 CATHERINE **HAMILTON** New 220 627088 **JAC** 91 Store NaN York **GROCERY GROCERY** PL**CORP** 3011 New **BROADWAY** 224 621015 **GROCERY** C TOWN 3320 Store JAC NaN York INC KING MEAT New 225 708796 JAC **SHOP FAIR** 3871 **BROADWAY** NaN Store **CORP** York 91 rows × 18 columns In [158... # See convenience stores food_retailers_d7_df.loc[food_retailers_d7_df["grocery"] == False] DBA.Name Street.Number Street.Name Address.Line.2 Ad Out[158. County License.Number Operation.Type Establishment.Type **Entity.Name DOLLAR DOLLAR AMSTERDAM** New 0 627413 VARIETY **VARIETY** 762 NaN Store York **AVE** STORE INC STORE **GALAXY** New **GALAXY** 1 717156 **GOURMET BROADWAY JAC** 3778 NaN Store York **GOURMET** INC **FINEST FINEST AMSTERDAM** New 3 739966 **JAC COLUMBIA COLUMBIA** Store NaN York **AVE** DELI **DELI INC SANTIAGO SANTIAGO** New 4 628824 **JAC DELI FOOD** 3415 **BROADWAY** NaN Store **DELI FOOD** York **CORP ABC MART** New 6 725281 Α ABC MART 3505 **BROADWAY** NaN Store York NC **NEW DELI 1** New 227 741584 **JAC NEW DELI 1** 3660 **BROADWAY** Store NaN York **CORP UPPER UPPER AMSTERDAM** New 228 748699 1728 Store MANHATTAN **MANHATTAN** NaN **AVE** York PHARMA INC **PHARMACY** DUANE DUANE New 229 626558 **BROADWAY** NaN Store READE ETAL 2864 **READE 14225** York **PTRS JOES JOES COLUMBUS** New **GOURMET** 230 **JAC GOURMET** 920 721538 Store NaN **AVENUE** York **DELI INC DELI** 1967 99 New CENTS & 99 CENTS & 231 627132 AMSTERDAM Store Α 1967 NaN York **DISCOUNT** DISCOUNT INC 141 rows × 18 columns In [159... # Save csv food_retailers_d7_df.to_csv("updated_food_retailers_d7.csv") 3. Formatting for Mapbox Combined fast food, convenience store, and grocery store data: https://drive.google.com/file/d/1NeeFnY3lbDGXIV8DIiXqkxn77iptmzB6/view?usp=sharing In [160... combined_df = pd.read_csv("./combined_food.csv") combined df Out[160... **DBA Address** latitude longitude retailer_type 0 **DOLLAR VARIETY STORE** 762 AMSTERDAM AVE 40.794936 -73.969728 convenience 1 **GALAXY GOURMET** 3778 BROADWAY 40.833942 -73.944853 convenience 2 FINEST COLUMBIA DELI 945 AMSTERDAM AVE 40.800718 -73.965524 convenience 3 SANTIAGO DELI FOOD 3415 BROADWAY 40.822390 -73.953531 convenience 4 3505 BROADWAY **ABC MART** 40.825393 -73.951326 convenience 900 RIVERSIDE DRIVE DYLANS DELICIOUS DELI G 257 40.838297 -73.946282 grocery 2892 BROADWAY 258 MILANO MARKET 40.805828 -73.965361 grocery 259 91 CATHERINE GROCERY 91 HAMILTON PL 40.823238 -73.950316 grocery **C TOWN** 3320 BROADWAY 40.819700 -73.955240 260 grocery 261 SHOP FAIR 3871 BROADWAY 40.836855 -73.942971 grocery 262 rows × 5 columns We need to make the DBA and Address columns NOT all caps, and follow Mapbox formatting guidelines so we can map our data In [161... dba = combined_df["DBA"].str.title() combined_df["DBA"] = dba In [162... # Reformat addresses address = combined df["Address"].str.title().to list() new addresses = [] for a in address: a = a.replace("Ave", "Ave.").replace("Street", "St.").replace("St", "St.") a = a.replace("Avenue", "Ave.") new addresses.append(a) combined_df["Address"] = new_addresses In [163... # Rename columns combined_df = combined_df.rename(columns={"DBA" : "title", "Address" : "address"}) In [164... # Give more descriptive strings for retailer type retailer_type_list = combined_df["retailer_type"].to_list() new retailers = [] for r in retailer type list: if r == "convenience": new retailers.append("Convenience stores and delis") elif r == "grocery": new retailers.append("Grocery stores") new retailers.append("Top 50 American fast-food chains") combined_df["retailer_type"] = new_retailers In [165... combined_df latitude longitude Out[165... title address retailer_type 0 Dollar Variety Store 762 Amsterdam Ave. 40.794936 -73.969728 Convenience stores and delis 1 **Galaxy Gourmet** 3778 Broadway 40.833942 -73.944853 Convenience stores and delis Finest Columbia Deli 945 Amsterdam Ave. 40.800718 -73.965524 Convenience stores and delis 2 3 Santiago Deli Food 3415 Broadway 40.822390 -73.953531 Convenience stores and delis 3505 Broadway 40.825393 -73.951326 Convenience stores and delis 4 Abc Mart **257** Dylans Delicious Deli G 900 Riverside Drive 40.838297 -73.946282 Grocery stores Milano Market 2892 Broadway 40.805828 -73.965361 258 Grocery stores 259 91 Catherine Grocery 91 Hamilton Pl 40.823238 -73.950316 Grocery stores 260 C Town 3320 Broadway 40.819700 -73.955240 Grocery stores 261 Shop Fair 3871 Broadway 40.836855 -73.942971 Grocery stores 262 rows × 5 columns The remaining formatting will require some manual spot-checking. Download so we can do just that: In [167... # Save csv

combined_df.to_csv("combined_food_before_manual_cleaning.csv")