**Task 1**: Perform basic analysis of Yelp data and provide results for at least completeness, frequency, dummy or empty values.

Ovaj zadatak ću započeti unošenjem podataka u Python koristeći biblioteku pandas.

import pandas as pd  
  
data = pd.read\_csv("C:\Users\Edna i Anesa\Desktop\yelp\_business.csv", sep ='|', low\_memory=False)

1. **Frekvencija**

Nakon određivanja frekvencije svake kolone, najrelevantnije cu priložiti u izvještaju.

Svaki element iz kolone *business\_id* ima frekvenciju 1, što je očekivano jer je to atribut koji mora biti unikatan za svaki posao.

U sljedećoj tabeli su prikazane frekvencije država. Kako su svi podaci iz SAD-a, a SAD čini 50 država, vidimo da se pojavljuje 36 različitih od 50.

|  |  |
| --- | --- |
| AZ 56686 | NV 36312 |
| ON 33412 | NC 14720 |
| OH 14697 | PA 11216 |
| QC 9219 | AB 8012 |
| WI 5154 | IL 1932 |
| SC 1162 | NY 22 |
| CA 19 | TX 6 |
| XGM 4 | FL 4 |
| AL 3 | WA 3 |
| CT 3 | VT 2 |
| VA 2 | AK 2 |
| NE 2 | GA 2 |
| XWY 2 | NJ 1 |
| NM 1 | BAS 1 |
| TN 1 | XGL 1 |
| CON 1 | AR 1 |
| DUR 1 | BC 1 |
| UT 1 | DOW 1 |

U sljedećoj tabeli se nalaze frekvencije pojavljivanja različitih gradova. U nastavku analize ovi podaci će biti korisni u kombinaciji sa frekvencijom država.

|  |  |  |
| --- | --- | --- |
| "Las Vegas" 29370 | Toronto 18906 | Phoenix 18766 |
| Charlotte 9509 | Scottsdale 8837 | Calgary 7736 |
| Pittsburgh 7017 | Montréal 6449 | Mesa 6080 |
| Henderson 4892 | Tempe 4550 | Chandler 4309 |
| Cleveland 3605 | Glendale 3543 | Madison 3494 |
| Gilbert 3462 | Mississauga 3112 | Peoria 1919 |
| Markham 1766 | "North Las Vegas" 1548 | Champaign 1262 |
| "North York" 1205 | Surprise 1148 | Scarborough 1103 |
| "Richmond Hill" 1026 | Brampton 1002 | Concord 990 |
| Vaughan 919 | Goodyear 855 | Etobicoke 793 |
| Matthews 749 | Avondale 681 | "Fort Mill" 627 |
| Huntersville 625 | Lakewood 519 | Gastonia 480 |
| Cornelius 474 | Mentor 472 | Akron 465 |
| Urbana 411 | "Cave Creek" 402 | Monroeville 400 |
| "North Olmsted" 396 | Thornhill 396 | Laval 394 |
| Westlake 389 | Strongsville 379 | Oakville 362 |
| Middleton 360 | Aurora 356 | "Cuyahoga Falls" 354 |
| Newmarket 352 | Pineville 349 | Medina 348 |
| "Fountain Hills" 345 | Pickering 326 | "Rock Hill" 311 |
| Ajax 303 | "Indian Trail" 301 | Parma 290 |
| Beachwood 289 | Wexford 280 | Whitby 276 |
| "Boulder City" 269 | "Sun City" 258 | Willoughby 257 |
| Kent 254 | "Sun Prairie" 246 | "Rocky River" 245 |
| Woodbridge 241 | Stow 234 | Avon 230 |
| "Litchfield Park" 222 | Solon 212 | "Cleveland Heights" 212 |
| Elyria 208 | Hudson 203 | Fitchburg 201 |
| "Bethel Park" 200 | Verona 198 | "Chagrin Falls" 196 |
| Bridgeville 191 | Brossard 177 | Brunswick 176 |
| Monroe 175 | Coraopolis 170 | Waxhaw 170 |
| Airdrie 168 | Canonsburg 163 | "Mayfield Heights" 162 |
| "East York" 161 | Belmont 160 | Fairlawn 157 |
| "North Royalton" 157 | Verdun 153 | Montreal 153 |
| Tolleson 147 | "West Mifflin" 145 | Independence 143 |
| "Middleburg Heights" 139 | Monona 137 | Streetsboro 137 |
| "Paradise Valley" 136 | Saint-Laurent 136 | Painesville 132 |
| Twinsburg 130 | Laveen 127 | Harrisburg 123 |
| "Avon Lake" 122 | "El Mirage" 122 | Berea 121 |
| Homestead 119 | Lyndhurst 118 | "Broadview Heights" 118 |
| York 116 | Denver 115 | Stoughton 115 |
| Sewickley 115 | Bedford 114 | Lorain 113 |
| Gibsonia 113 | Carnegie 112 | Murrysville 111 |
| "Mint Hill" 110 | Davidson 108 | "Fairview Park" 106 |
| Euclid 106 | Amherst 105 | "Sun City West" 104 |
| Pointe-Claire 102 | Carefree 100 | "Queen Creek" 100 |
| Kannapolis 100 | Macedonia 99 | Dorval 97 |
| "North Ridgeville" 97 | Westmount 97 | "Allison Park" 96 |
| McMurray 93 | Northfield 91 | Chardon 87 |
| Dollard-des-Ormeaux 87 | Waunakee 86 | "New Kensington" 81 |
| Brecksville 81 | "South Euclid" 79 | Savoy 79 |
| Brooklyn 78 | Longueuil 78 | Oregon 76 |
| "Indian Land" 75 | Irwin 74 | Eastlake 74 |
| Lasalle 73 | Wickliffe 72 | Buckeye 69 |
| Willowick 69 | "Parma Heights" 69 | Chesterland 67 |
| Vaudreuil-Dorion 67 | "Shaker Heights" 66 | Youngtown 65 |
| "North Versailles" 63 | Woodmere 60 | McKeesport 60 |
| Stouffville 60 | Maple 59 | Tarentum 55 |
| "Brook Park" 55 | Oakmont 54 | Lachine 54 |
| "Lake Wylie" 53 | "McKees Rocks" 53 | Tallmadge 53 |
| "Mount Horeb" 52 | "Maple Heights" 52 | Glenshaw 51 |
| "Garfield Heights" 50 | Pierrefonds 50 | Rantoul 50 |
| Saint-Sauveur 50 | Unionville 49 | "Willoughby Hills" 48 |
| Oakdale 48 | "Mount Holly" 48 | "Moon Township" 47 |
| Kirkland 47 | "Olmsted Falls" 47 | Whitchurch-Stouffville 47 |
| "Bay Village" 46 | Clover 46 | "Warrensville Heights" 45 |
| Richfield 45 | Mooresville 45 | Bolton 44 |
| "University Heights" 42 | Mahomet 42 | Bellevue 41 |
| Outremont 38 | "North Huntingdon" 38 | "Highland Heights" 38 |
| Copley 37 | "Cottage Grove" 37 | "Mc Kees Rocks" 37 |
| "Pleasant Hills" 36 | Monticello 36 | Ambridge 36 |
| Elizabeth 35 | "Seven Hills" 34 | "White Oak" 34 |
| Bradford 33 | "Penn Hills" 33 | Boucherville 33 |
| Caledon 33 | Georgetown 32 | Saint-Leonard 32 |
| Mont-Royal 31 | Munhall 31 | "Mc Farland" 31 |
| Saint-Jean-sur-Richelieu 31 | "Tega Cay" 31 | "Sun Lakes" 31 |
| Chestermere 31 | "Halton Hills" 30 | Peninsula 29 |
| Cheswick 29 | "Sheffield Village" 29 | McFarland 28 |
| Dollard-Des-Ormeaux 28 | Middlefield 28 | Dallas 28 |
| "Richmond Heights" 27 | "Columbia Station" 27 | "East Gwillimbury" 27 |
| "Apache Junction" 26 | Burton 26 | Wadsworth 26 |
| "N Las Vegas" 26 | "De Forest" 26 | Anthem 26 |
| "Bedford Heights" 25 | Higley 25 | Guadalupe 25 |
| "Cross Plains" 24 | "Valley View" 24 | Paradise 24 |
| Saint-Hubert 23 | "Robinson Township" 22 | Stallings 22 |
| "Lower Burrell" 22 | Saint-Lambert 22 | "Wesley Chapel" 21 |
| Finleyville 21 | Imperial 21 | "La Prairie" 21 |
| "Rocky View" 20 | "Munroe Falls" 20 | "King City" 20 |
| Anjou 20 | "Greenfield Park" 20 | Midland 20 |
| Plum 19 | LaSalle 19 | "Pepper Pike" 19 |
| "Bradford West Gwillimbury" 19 | Hinckley 19 | Newbury 19 |
| Saint-Eustache 19 | Waddell 19 | Millvale 19 |
| "West Homestead" 19 | Ahwatukee 19 | Stanley 19 |
| "LAS VEGAS" 19 | Mayfield 18 | "Turtle Creek" 18 |
| Aspinwall 18 | Kirtland 18 | DeForest 18 |
| "Mckees Rocks" 18 | Rosemère 18 | Saint-Jérôme 18 |
| Brentwood 18 | Sainte-Anne-de-Bellevue 17 | "Saint Joseph" 17 |
| Terrebonne 17 | Trafford 16 | Monongahela 16 |
| "West View" 16 | "Nellis AFB" 16 | Clairton 16 |
| "Valley City" 16 | Grafton 16 | Sainte-Julie 16 |
| Brookpark 16 | Windsor 15 | "Laveen Village" 15 |
| Mantua 15 | "Mount Lebanon" 15 | "Castle Shannon" 15 |
| Boisbriand 15 | Ravenna 14 | "Las vegas" 14 |
| "Jefferson Hills" 14 | Milton 14 | Belleville 14 |
| Lowell 14 | "Sheffield Lake" 14 | "Hampton Township" 13 |
| Wilkinsburg 13 | Springdale 13 | Leetsdale 13 |
| Chambly 13 | Châteauguay 13 | Locust 13 |
| Crafton 12 | Orange 12 | "N. Las Vegas" 12 |
| "Oakwood Village" 12 | Cramerton 12 | Heidelberg 12 |
| "Blue Diamond" 12 | Etna 12 | "Mc Murray" 12 |
| Robinson 11 | Sainte-Thérèse 11 | Chateauguay 11 |
| "South Park" 11 | Beloeil 11 | Weddington 11 |
| Saint-Hyacinthe 11 | "Fort McDowell" 11 | Balzac 11 |
| Litchfield 11 | Sheffield 11 | Wilmerding 11 |
| Summerlin 11 | Beaconsfield 11 | Braddock 11 |
| Blainville 11 | Schomberg 10 | Champlain 10 |
| Uxbridge 10 | "Orange Village" 10 | "Mayfield Village" 10 |
| Nobleton 10 | Dravosburg 10 | Norton 10 |
| "Moreland Hills" 10 | Brooklin 10 | Deforest 10 |
| Cecil 9 | McDonald 9 | "Upper Saint Clair" 9 |
| Swissvale 9 | "North Randall" 9 | Sainte-Adele 9 |
| Mont-Saint-Hilaire 9 | "Rocky View County" 9 | phoenix 9 |
| "Caledon East" 9 | "Spring Valley" 9 | Bainbridge 9 |
| L'Île-Perrot 9 | Kleinburg 9 | "Sagamore Hills" 9 |
| Saint-Jerome 9 | Mentor-on-the-Lake 9 | "Black Earth" 9 |
| Moon 9 | Sharpsburg 9 | Dormont 8 |
| Pittsburg 8 | Willowdale 8 | "Mt Lebanon" 8 |
| "South Park Township" 8 | Avalon 8 | "las vegas" 8 |
| Saint-Léonard 8 | Saint-Lazare 8 | HENDERSON 8 |
| Sainte-Anne-De-Bellevue 8 | Mcmurray 8 | Sainte-Adèle 8 |
| Mckeesport 8 | Rosemere 8 | "Ross Township" 7 |
| Venetia 7 | "Mt. Lebanon" 7 | "Cote Saint-Luc" 7 |
| "East Cleveland" 7 | Delson 7 | "Rouses Point" 7 |
| Enterprise 7 | Lachute 7 | Montreal-Nord 7 |
| "McCandless Township" 7 | Hiram 7 | CHARLOTTE 7 |
| "La Salle" 7 | "Harrison City" 7 | "Mc Donald" 7 |
| Joliette 6 | "East Pittsburgh" 6 | "Belle Vernon" 6 |
| Mirabel 6 | "Upper St. Clair" 6 | McCandless 6 |
| Morin-Heights 6 | PHOENIX 6 | MESA 6 |
| "Dollard-des Ormeaux" 6 | Blawnox 6 | "Middleburg Hts" 6 |
| Saint-Constant 6 | "Mentor On the" 6 | "Mentor On The Lake" 6 |
| Perry 6 | Ogden 6 | Glassport 6 |
| Vimont 6 | Dane 6 | "Olmsted Township" 6 |
| "Upper St Clair" 6 | Oka 6 | Duquesne 6 |
| Saint-Bruno-de-Montarville 6 | "Fairport Harbor" 6 | Herminie 5 |
| Pitcairn 5 | scottsdale 5 | Oberlin 5 |
| "Walton Hills" 5 | Pointe-Aux-Trembles 5 | Kahnawake 5 |
| Rexdale 5 | "Concord Township" 5 | Mascouche 5 |
| Bellvue 5 | Downtown 5 | L'ile-Perrot 5 |
| "Las Vegas" 5 | Harmarville 5 | "Mentor On the Lake" 5 |
| TEMPE 5 | Beeton 5 | "Mount Albert" 5 |
| St-Leonard 5 | "East McKeesport" 5 | "Rio Verde" 5 |
| Export 5 | King 5 | Novelty 5 |
| Bath 5 | Presto 5 | Lagrange 5 |
| Gormley 5 | "Holland Landing" 4 | Sainte-Catherine 4 |
| "Gates Mills" 4 | Missisauga 4 | "New Eagle" 4 |
| Library 4 | LaGrange 4 | Montréal-Ouest 4 |
| Mont-Saint-Grégoire 4 | Shadyside 4 | Russellton 4 |
| Montreal-Ouest 4 | McAdenville 4 | "Northfield Center" 4 |
| Homer 4 | "Nellis Air Force Base" 4 | "Rocky View No. 44" 4 |
| "Las Vegas " 4 | Sainte-Genevieve 4 | Piedmont 4 |
| "Concord Mills" 4 | Clark 4 | Rigaud 4 |
| Lawrenceville 4 | "Port Credit" 4 | Oakwood 4 |
| Saint-Bruno 4 | "Downtown Toronto" 4 | Arnold 4 |
| Oakland 4 | "South Las Vegas" 4 | "Don Mills" 4 |
| Tottenham 4 | Montréal-Nord 3 | Salaberry-de-Valleyfield 3 |
| "Fort Mcdowell" 3 | Weston 3 | Wildwood 3 |
| McMurry 3 | charlotte 3 | Lancaster 3 |
| "Nellis Afb" 3 | Emsworth 3 | henderson 3 |
| Bratenahl 3 | mesa 3 | "N Ridgeville" 3 |
| Scottdale 3 | Streetsville 3 | "East Liberty" 3 |
| "Division No. 6" 3 | Salaberry-De-Valleyfield 3 | "Squirrel Hill" 3 |
| Montreal-West 3 | Saint-Bernard-de-Lacolle 3 | Mcfarland 3 |
| "Kennedy Township" 3 | "Northwest Calgary" 3 | Candiac 3 |
| St-Bruno-de-Montarville 3 | "Warrensville Hts" 3 | "East Mc Keesport" 3 |
| Clarkson 3 | Marvin 3 | Pheonix 3 |
| "Pointe Claire" 3 | SCOTTSDALE 3 | Repentigny 3 |
| "Medina Township" 3 | "Cuyahoga Fls" 3 | Sainte-Therese 3 |
| Paoli 3 | Marshall 3 | "Forest Hills" 3 |
| St-Jerome 3 | Creighton 3 | "Southeast Calgary" 3 |
| "Brooklyn Heights" 3 | Cote-Saint-Luc 3 | Bois-des-Filion 3 |
| Philo 3 | toronto 3 | "Highland Hills" 3 |
| "Côte Saint-Luc" 3 | "Richmond Hts" 3 | Sidney 3 |
| Rawdon 3 | Noyan 2 | "San Tan Valley" 2 |
| "Ft. Mill" 2 | Erie 2 | "St Joseph" 2 |
| Mercier 2 | "Central City Village" 2 | Gifford 2 |
| Downsview 2 | Sutersville 2 | Malton 2 |
| "Northfield Center Township" 2 | Warrensville 2 | "Woodmere Village" 2 |
| "St Leonard" 2 | Roxboro 2 | "Peters Township" 2 |
| Montrose 2 | Tuscola 2 | Etobiicoke 2 |
| glendale 2 | Ormstown 2 | "Dollard-Des Ormeaux" 2 |
| Coteau-du-Lac 2 | Saint-Marc-sur-Richelieu 2 | "Painesville Township" 2 |
| Hemmingford 2 | Sainte-Anne-des-Plaines 2 | Mathews 2 |
| Saint-Philippe 2 | "Broadview Hts" 2 | "Shaler Township" 2 |
| Ingomar 2 | "North Olmstead" 2 | "Las Vegas, NV" 2 |
| "New River" 2 | Rockyview 2 | "Eighty Four" 2 |
| Scarbrough 2 | "Durham Regional Municipality" 2 | Bloomfield 2 |
| Pincourt 2 | "Moon Twp" 2 | CONCORD 2 |
| Ashburn 2 | Beauharnois 2 | Cooksville 2 |
| "Cuyahoga Heights" 2 | Huntsburg 2 | "Shaker Hts" 2 |
| Napierville 2 | "Garfield Hts" 2 | LasVegas 2 |
| "Saint Laurent" 2 | Montral 2 | Cuddy 2 |
| "Mt. Holly" 2 | Kettleby 2 | L'ile-Bizard 2 |
| "Oak Ridges" 2 | Communauté-Urbaine-de-Montréal 2 | "North Hills" 2 |
| "York Regional Municipality" 2 | "Lake Park" 2 | Pointe-Calumet 2 |
| TORONTO 2 | "Deux-Montagnes Regional County Municipality" 2 | Lowesville 2 |
| Brookline 2 | "Concord Twp" 2 | "South Hills" 2 |
| Tremont 2 | "North Huntington" 2 | Ross 2 |
| Côte-Saint-Luc 2 | "Charlotte " 2 | "Phoenix Metro Area" 2 |
| "Grand River" 2 | "Plum Boro" 2 | "Phoenix, AZ" 2 |
| Sturgeon 2 | "Pittsburgh " 2 | Saint-Hippolyte 2 |
| "Old Town Scottsdale" 2 | Farnham 2 | Edgewood 2 |
| Val-Morin 2 | Monoroeville 2 | GILBERT 2 |
| Baldwin 2 | "West Elizabeth" 2 | Saint-Charles-Borromee 2 |
| "La Vegas" 2 | Barberton 2 | Rankin 2 |
| "Mt Holly" 2 | Morgan 2 | Saint-Sauveur-des-Monts 2 |
| "Sauk City" 2 | Leeds 2 | "Boston Heights" 2 |
| Lachenaie 2 | Saint-Jean-Sur-Richelieu 2 | "Fort mill" 2 |
| "North Braddock" 2 | Geauga 2 | Prévost 2 |
| Goodwood 2 | "Mayfield Hts" 2 | Vaudreuil 2 |
| "N. Olmsted" 2 | Edmonton 2 | "Shorewood Hills" 2 |
| Fisher 2 | "Indian land" 2 | "Silver Lake" 2 |
| "Franklin Park" 2 | "Summerlin South" 2 | Rillton 2 |
| Baie-d'Urfé 2 | "Los Angeles" 2 | Deerfield 2 |
| "Neville Island" 2 | L'Île-Bizard 2 | "East Mississauga" 2 |
| Whitney 1 | L'Assomption 1 | Edgemont 1 |
| "North Carolina" 1 | calgary 1 | lyndhurst 1 |
| "Pgh Int Arprt" 1 | "Chomedey, Laval" 1 | Alberta 1 |
| Alburgh 1 | Mississauge 1 | McMasterville 1 |
| "Troy Township" 1 | Joliet 1 | "N. Ridgeville" 1 |
| "Henderson Nevada" 1 | "South Russell" 1 | "boulder city" 1 |
| Indianola 1 | Banksville 1 | Palgrave 1 |
| "Monterey Park" 1 | "Lake Norman" 1 | Down 1 |
| "Auburn Twp" 1 | "North york" 1 | GIlbert 1 |
| Char 1 | University 1 | Montréal-Est 1 |
| Whiitby 1 | kirtland 1 | Hampstead 1 |
| "N E Las Vegas" 1 | AGINCOURT 1 | Chatauguay 1 |
| "Avon lake" 1 | Hamilton 1 | Union 1 |
| Sainte-Geneviève 1 | St-Lazare 1 | montreal 1 |
| "Westworld Scottsdale" 1 | Tampa 1 | Mississuaga 1 |
| "North Las Vegas " 1 | "W Spring Valley" 1 | "Las Vegas East" 1 |
| "Auburn Township" 1 | "Nottingham Township" 1 | "Mint Hill" 1 |
| Crescent 1 | Bury 1 | MACEDONIA 1 |
| Mississauaga 1 | Closter 1 | "Rockey View" 1 |
| Tornto 1 | Sainte-Rose 1 | Berry 1 |
| Deux-Montagnes 1 | Mentor-On-The-Lake 1 | Sainte-Marguerite-Esterel 1 |
| Blakeney 1 | Ste-Dorothée 1 | "Lower Lawrenceville" 1 |
| "Lower burrell" 1 | Sainte-Dorothée 1 | "San Diego" 1 |
| Seattle 1 | CARSON 1 | ​Chandler 1 |
| "CAVE CREEK" 1 | "Otterburn Park" 1 | "Green Tree" 1 |
| "Regent Square" 1 | Chander 1 | "Mt. Horeb" 1 |
| "NELLIS AFB" 1 | "Las Vegas Valley" 1 | GOODYEAR 1 |
| Bradfordwoods 1 | "Richmond Height" 1 | Harrisburg, 1 |
| Etibicoke 1 | Baie-d'Urfe 1 | "Lakewood, Oh" 1 |
| "Shady Side" 1 | Richmonhill 1 | Ste-Rose 1 |
| "Cleveland Heigh" 1 | Queensville 1 | "De Winton" 1 |
| "Columbia Sta" 1 | Saint-laurent 1 | "Las Vegas Nv" 1 |
| Twinsburgh 1 | Milford 1 | Decatur 1 |
| Evergreen 1 | Ballantyne 1 | "East Credit" 1 |
| "New York" 1 | Saint-Pie 1 | "Quartier des Spectacles" 1 |
| Venise-en-Québec 1 | tempe 1 | Arrowhead 1 |
| Pepperpike 1 | "Saint - Hyacinthe" 1 | "Las Vegas," 1 |
| "Las Vegass" 1 | Midnapore 1 | Greensboro 1 |
| "Villa Grove" 1 | "South Amherst" 1 | Phoniex 1 |
| "Caledon Village" 1 | "Stetson Valley" 1 | Phoneix 1 |
| "Le Sud-Ouest" 1 | "Brunswick Hills" 1 | "Tempe " 1 |
| Sunrise 1 | Puyallup 1 | "Chargrin Falls" 1 |
| Manchester 1 | Westview 1 | Sacramento 1 |
| "C Las Vegas" 1 | Saintt-Bruno-de-Montarville 1 | Antioch 1 |
| "Bedford Hts." 1 | Phoenx 1 | "Mount Oliver" 1 |
| "Cleveland Height" 1 | "Las Vegas NV" 1 | "Pheonix AZ" 1 |
| "East Mesa" 1 | Notre-Dame-De-L'ile-Perrot 1 | Canonsburd 1 |
| Frazer 1 | Huntingdon 1 | "110 Las Vegas" 1 |
| Central 1 | Aliquippa 1 | "Sage Hill" 1 |
| SCARBOROUGH 1 | Reminderville 1 | Oldham 1 |
| "Old Port of Montreal" 1 | Sainte-Thérèse-de-Blainville 1 | "Le Haut-Saint-Laurent Regional County Municipality" 1 |
| Phoeniix 1 | "las Vegas" 1 | "Cleveland Hghts." 1 |
| "Wilkins Township" 1 | MATTHEWS 1 | Maricopa 1 |
| "Les Cèdres" 1 | Montgomery 1 | "Warrensville Hts." 1 |
| "Murrysville (Monroeville)" 1 | "St. Léonard" 1 | "Wilkens Township" 1 |
| "Cranberry Twp" 1 | "Penn Hills Township" 1 | "Vaughan Mills" 1 |
| Pasadena 1 | "Green Valley" 1 | "Central Oakland" 1 |
| Broadlands 1 | Allegheny 1 | "St. Sauveur" 1 |
| "Northeast Calgary" 1 | Montreal-Est 1 | "North Olmsted," 1 |
| Saint-Isidore-de-Laprairie 1 | Lacolle 1 | Inglewood 1 |
| Varennes 1 | "Elizabeth Township" 1 | "Mayfield Heights (Cleveland)" 1 |
| "Glen Williams" 1 | "Fox Chapel" 1 | Toronto-West 1 |
| "Henderson and Las vegas" 1 | "Strip District" 1 | "Cave Creek Road" 1 |
| "Richland Township" 1 | "River Drive Park" 1 | "Beachwood " 1 |
| Iberville 1 | Québec 1 | "Fountain Hls" 1 |
| Temecula 1 | "Woodbridge (Vaughan)" 1 | Pennsylvania 1 |
| "Queen creek" 1 | "Sun CIty" 1 | Montréal-West 1 |
| "Mccandless Township" 1 | "Ville Saint Laurent" 1 | Farmington 1 |
| "SUN PRAIRIE" 1 | Charlemagne 1 | "Saint Leonard" 1 |
| "Side Slopes" 1 | Montéal 1 | Schaumburg 1 |
| "Laval, Pont Viau" 1 | Dublin 1 | "S Concord" 1 |
| Saint-Basile-le-Grand 1 | Moseley 1 | Scottadale 1 |
| "Moon Twp." 1 | "Morgan Hill" 1 | "Lake Las Vegas" 1 |
| "Highland Hts" 1 | Tolono 1 | "St. Leonard" 1 |
| Chandler-Gilbert 1 | Montreal-Quest 1 | Whtiby 1 |
| Oakridges 1 | "Central Henderson" 1 | "Les Coteaux" 1 |
| Saint-Henri 1 | Godmanchester 1 | "Centre Island" 1 |
| Greenway 1 | "QUEEN CREEK" 1 | "NORTH YORK" 1 |
| Snowflake 1 | "Township of Concord" 1 | "Toronto-North York" 1 |
| SURPRISE 1 | Warrendale 1 | solon 1 |
| Peninsul 1 | "Regional Municipality of York" 1 | springdale 1 |
| "Ben Avon" 1 | Rostraver 1 | Estérel 1 |
| "FORT MILL" 1 | Thornhil 1 | St-Laurent 1 |
| "New Tecumseth" 1 | "Toronto Division" 1 | "N W Las Vegas" 1 |
| "Sunrise Manor" 1 | Point-Claire 1 | Scarobrough 1 |
| Toronto-Etobicoke 1 | "West Lake" 1 | Scotsdale 1 |
| "Red Rock" 1 | "St. Jean Sur Richelieu" 1 | "City Center" 1 |
| Allentown 1 | "Apache Trail" 1 | Sainte-Marguerite-du-lac-Masson 1 |
| Garrettsville 1 | Westport 1 | Mcdonald 1 |
| Austin 1 | "Ville Mont-Royal" 1 | PEORIA 1 |
| "Church Cove" 1 | "University Hts " 1 | TOLLESON 1 |
| Nationwide 1 | St.Pittsburgh 1 | Fitchburgh 1 |
| Claremont 1 | Gelndale 1 | Sainte-thérèse 1 |
| "Lake Mead" 1 | "Bond Head" 1 | "St. Joseph" 1 |
| Scotesdale 1 | "Las Vegas Nevada" 1 | "Île des Soeurs" 1 |
| Somerton 1 | "Urbana, Illinois" 1 | Mississaugua 1 |
| "Grand Canyon" 1 | "SW Calgary" 1 | Notre-Dame-de-l'Ile-Perrot 1 |
| Saint-Pierre-de-Véronne-à-Pike-River 1 | Lavaltrie 1 | Dewey 1 |
| Pahrump 1 | Valleywide 1 | peoria 1 |
| "Huntington Beach" 1 | "St-Benoît de Mirabel" 1 | "Rock Hill SC" 1 |
| "Mesa Arizona" 1 | Mississagua 1 | Spartanburg 1 |
| Montrèal 1 | Pine 1 | "Bainbridge Township" 1 |
| "New Glarus" 1 | "Rockyview County" 1 | Mattews 1 |
| "West Toronto" 1 | "W Henderson" 1 | "North Toronto" 1 |
| charlottte 1 | "Medicine Hat" 1 | "The Promenade" 1 |
| "Litchfield park" 1 | "North Phoenix" 1 | CALGARY 1 |
| "Rainbow Valley" 1 | "Middleburg Hts." 1 | Lufkin 1 |
| Oshawa 1 | "Metro Phoenix" 1 | "Mesa AZ" 1 |
| Agincourt 1 | "Catawba Springs" 1 | verdun 1 |
| "RIchmond Hill" 1 | "Cuyohoga Falls" 1 | "Saint-Bernard de Lacolle" 1 |
| Elrama 1 | "Toronto Scarborough" 1 | "Red Mountain" 1 |
| Lynhurst 1 | "Black Rock City" 1 | Columbus 1 |
| "Sandy Ridge" 1 | oakville 1 | "Grove City" 1 |
| "Henderson (Green Valley)" 1 | "San Tan" 1 | Etobicoke, 1 |
| moon 1 | Zephyr 1 | "Los Vegas" 1 |
| Chattanooga 1 | Mooers 1 | Scottsale 1 |
| "Rocky river" 1 | Bakerstown 1 | "Pickering Ajax Whitby" 1 |
| "Fort Mill" 1 | Sainte-Marguerite-du-Lac-Masson 1 | Mcknight 1 |
| Yorkdale 1 | "Spring Hill City View" 1 | Boston 1 |
| "Port Vue" 1 | Tucson 1 | AVON 1 |
| "Highland Park" 1 | "Hyland Heights" 1 | "Oakwood village" 1 |
| Leslieville 1 | "Mt. Washington" 1 | "North York" 1 |
| "Buena Vista" 1 | Brownsburg-Chatham 1 | "Surprise Crossing" 1 |
| "Cranberry Township" 1 | "cave creek" 1 | Promenade 1 |
| Median 1 | "Estrella Village" 1 | Ansnorveldt 1 |
| Beltline 1 | Darlington 1 | "Paw Creek" 1 |
| Ahwahtukee 1 | Uptown 1 | Auburn 1 |
| "Mayfield Heights." 1 | Montville 1 | ETOBICOKE 1 |
| Ivesdale 1 | Nevada 1 | Middleburg 1 |
| "NORTH LAS VEGAS" 1 | "East Hawkesbury" 1 | GLENDALE 1 |
| DENVER 1 | AZ 1 | LAVEEN 1 |
| Thomasboro 1 | "Robinson Twp." 1 | Harwick 1 |
| Crouse 1 | "East Gwilimbury" 1 | THORNHILL 1 |
| Fabreville 1 | "Phoenix AZ" 1 | "Russell Twp" 1 |
| "Brooklyn Hts." 1 | Bentleyville 1 | Suprise 1 |
| "North Charlotte" 1 | "Harmar Township" 1 | "Ahwatukee Foothills Village" 1 |
| Phx 1 | Ingram 1 | Sunnyslope 1 |
| Pointe-aux-Trembles 1 | Harrisbug 1 | "Cleveland Hts." 1 |
| "O'hara Township" 1 | Sedona 1 | "Las Vegas & Henderson" 1 |
| Monreoville 1 | Glbert 1 | Warrenville 1 |
| Sale 1 | MADISON 1 | "North of Brampton" 1 |
| Rougemont 1 | McKnight 1 | "Phoenix " 1 |
| Phoenix, 1 | "Centennial Hills" 1 | St-Clet 1 |
| Vaughn 1 | "Blue Mounds" 1 | L'assomption 1 |
| "Indian Land," 1 | "East Calgary" 1 | "University Ht" 1 |
| "Markham, ON" 1 | "South Charlotte" 1 | Yorkville 1 |
| Pittsburch 1 | Mansfield 1 | Wellington 1 |
| Leaside 1 | "Perry Twp" 1 | Hendersonville 1 |
| Citibank 1 | "Southside Flats" 1 | Suncity 1 |
| NorthYork 1 | NEWMARKET 1 | "Montréal (Québec)" 1 |
| "N Solon" 1 | Fithian 1 | Sharon 1 |
| "N. Randall" 1 | Schottsdale 1 | markham 1 |
| CAREFREE 1 | "Glendale Az" 1 | Chateau 1 |
| "Mayfield Hts." 1 | Nellis 1 | Burlington 1 |
| "Old Brooklyn" 1 | St-Jean-sur-Richelieu 1 | Saint-Roch-de-l'Achigan 1 |
| Midway 1 | "Richmond Hil" 1 | "Mc Adenville" 1 |
| Saint-Basile-Le-Grand 1 | "Mont St-hilaire" 1 | "Maple Grove" 1 |
| "Rural Ridge" 1 | Norval 1 | Omaha 1 |
| etobicoke 1 | "Cleveland, OH" 1 | Hnederson 1 |
| Lindale 1 | Belœil 1 | Ranlo 1 |
| Sainte-Madeleine 1 | London 1 | "Parma mid birth" 1 |
| "south Euclid" 1 | Ste-Therese-de-Blainville 1 | Valencia 1 |
| McKeesRocks 1 | "Cave creek" 1 | "Phoenix Valley" 1 |
| "Buckeye - Shaker" 1 | "North Ridgeville" 1 | Henderston 1 |
| "Thorncliffe Park" 1 | Vieux-Montréal 1 | Arizona 1 |
| "Las Vergas" 1 | Waterloo 1 | "Newburgh Heights" 1 |
| Ange-Gardien 1 | Shaler 1 | "PARMA HEIGHTS" 1 |
| "Central City" 1 | Vegas 1 | Lawrence 1 |
| CLEVELAND 1 | "Mount Washington" 1 | "North Scottsdale" 1 |
| Draper 1 | "Bedford Hts" 1 | Sainte-Marthe 1 |
| "North Strabane Township" 1 | "West Montréal" 1 | "East Ajax" 1 |
| Lodi 1 | Henryville 1 | "Moon TWP" 1 |
| Chertsey 1 | "Laval, Ste Dorothee" 1 | "Old Scottsdale" 1 |
| GASTONIA 1 | "Scottsdale, AZ" 1 | "Cote Saint Luc" 1 |
| Northyork 1 | "Sun Praiie" 1 | Île-des-Soeurs 1 |
| "INDIAN LAND" 1 | Strongville 1 | "Desert Ridge" 1 |
| Alburg 1 | Clinton 1 | surprise 1 |
| Hendserson 1 | "De l'Eglise" 1 | MONROEVILLE 1 |

Prikazat ćemo također frekvenciju kolone *is\_open*.

|  |  |
| --- | --- |
| 1 | 158525 |
| 0 | 34084 |

U sljedećoj tabeli je prikazano koliko je poslova dobilo koju ocjenu, od 1 do 5, sa korakom 0.5.

|  |  |
| --- | --- |
| 4.0 | 35969 |
| 3.5 | 35008 |
| 5.0 | 28216 |
| 4.5 | 27301 |
| 3.0 | 25996 |
| 2.5 | 18843 |
| 2.0 | 11426 |
| 1.5 | 4976 |
| 1.0 | 4874 |

Ovo su neke bitnije određene frekvencije, sve se nalaze u prilogu.

1. **Null values**

Koristeći funkciju dataframe.isna().sum() za svaku kolonu određujemo broj podataka u toj koloni koji jesu null/NaN. Rezultati su u sljedećoj tabeli.

|  |  |
| --- | --- |
| address | 7682 |
| business\_id | 0 |
| categories | 482 |
| city | 1 |
| is\_open | 0 |
| latitude | 0 |
| longitude | 0 |
| name | 0 |
| postal\_code | 659 |
| review\_count | 0 |
| stars | 0 |
| state | 0 |
| geometry\_point | 0 |

1. **Completeness**

Koristeći funkciju dataframe.count() za svaku kolonu određujemo broj podataka u toj koloni koji nisu null/NaN. Rezultati su u sljedećoj tabeli.

|  |  |
| --- | --- |
| address | 184927 |
| business\_id | 192609 |
| categories | 192127 |
| city | 192608 |
| is\_open | 192609 |
| latitude | 192609 |
| longitude | 192609 |
| name | 192609 |
| postal\_code | 191950 |
| review\_count | 192609 |
| stars | 192609 |
| state | 192609 |
| geometry\_point | 192609 |

1. Dummy values

U sljedećoj tabeli je prikazano koliko se nalazi dummy vrijednosti u svakoj koloni.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | removed | nan | unknown | none | no name |
| Address | 0 | 0 | 0 | 0 | 0 |
| business\_id | 0 | 0 | 0 | 0 | 0 |
| Categories | 0 | 0 | 0 | 0 | 0 |
| City | 0 | 0 | 0 | 0 | 0 |
| Is\_open | 0 | 0 | 0 | 0 | 0 |
| latitude | 0 | 0 | 0 | 0 | 0 |
| Longitude | 0 | 0 | 0 | 0 | 0 |
| Name | 1 | 1 | 2 | 1 | 0 |
| postal\_code | 0 | 0 | 0 | 0 | 0 |
| review\_count | 0 | 0 | 0 | 0 | 0 |
| stars | 0 | 0 | 0 | 0 | 0 |
| state | 0 | 0 | 0 | 0 | 0 |
| geometry\_point | 0 | 0 | 0 | 0 | 0 |

Vidimo da se dummy vrijednosti pojavljuju jedino u koloni *name,* i zaključujemo da je odlično što podaci sadrže vrlo malo dummy vrijednosti.

**Task 2: Detect suspicious/invalid patterns and provide examples of these cases.**

U okviru ovog zadatka, odlučila sam odraditi sljedeće podzadatke:

* Provjeriti koliko poštanskih brojeva ne odgovara formatu SAD-a i izdvojiti ih
* Provjeriti koliko ima adresa koje ne odgovaraju očekivanom zapisu adrese, te ih izdvojiti
* Provjeriti koji gradovi ne pripadaju odgovarajućim državama
* Provjeriti odraničenja na koordinate
* Provjeriti da li adrese odgovaraju koordinatama