

Data Scientist Role Play: Profiling and Analyzing the Yelp Dataset Coursera Worksheet

This is a 2-part assignment. In the first part, you are asked a series of questions that will help you profile and understand the data just like a data scientist would. For this first part of the assignment, you will be assessed both on the correctness of your findings, as well as the code you used to arrive at your answer. You will be graded on how easy your code is to read, so remember to use proper formatting and comments where necessary.

In the second part of the assignment, you are asked to come up with your own inferences and analysis of the data for a particular research question you want to answer. You will be required to prepare the dataset for the analysis you choose to do. As with the first part, you will be graded, in part, on how easy your code is to read, so use proper formatting and comments to illustrate and communicate your intent as required.

For both parts of this assignment, use this "worksheet." It provides all the questions you are being asked, and your job will be to transfer your answers and SQL coding where indicated into this worksheet so that your peers can review your work. You should be able to use any Text Editor (Windows Notepad, Apple TextEdit, Notepad ++, Sublime Text, etc.) to copy and paste your answers. If you are going to use Word or some other page layout application, just be careful to make sure your answers and code are lined appropriately.

In this case, you may want to save as a PDF to ensure your formatting remains intact for you reviewer.

Part 1: Yelp Dataset Profiling and Understanding

1. Profile the data by finding the total number of records for each of the tables below:

- i. Attribute table = 10000

1	Select *
2	From Attribute
3	

Run

Reset

business_id	name	value
YDf95gJZaq05wvo7hTQbbQ	RestaurantsPriceRange2	2
YDf95gJZaq05wvo7hTQbbQ	BusinessParking	{"garage": false, "street": false, "valid
YDf95gJZaq05wvo7hTQbbQ	BikeParking	1
YDf95gJZaq05wvo7hTQbbQ	WheelchairAccessible	1
mLwM-h2YhXl2NCgdS84_Bw	GoodForMeal	{"dessert": false, "latenight": false, "l
mLwM-h2YhXl2NCgdS84_Bw	HasTV	0
mLwM-h2YhXl2NCgdS84_Bw	RestaurantsGoodForGroups	1
mLwM-h2YhXl2NCgdS84_Bw	NoiseLevel	average
mLwM-h2YhXl2NCgdS84_Bw	RestaurantsAttire	casual
mLwM-h2YhXl2NCgdS84_Bw	RestaurantsReservations	0
mLwM-h2YhXl2NCgdS84_Bw	OutdoorSeating	0
mLwM-h2YhXl2NCgdS84_Bw	BusinessAcceptsCreditCards	0
mLwM-h2YhXl2NCgdS84_Bw	RestaurantsPriceRange2	2
mLwM-h2YhXl2NCgdS84_Bw	RestaurantsDelivery	1
mLwM-h2YhXl2NCgdS84_Bw	Ambience	{"romantic": false, "intimate": false, "c
mLwM-h2YhXl2NCgdS84_Bw	RestaurantsTakeOut	1
mLwM-h2YhXl2NCgdS84_Bw	GoodForKids	1
v2WhjAB3PIBA8J8VxG3wEg	BusinessParking	{"garage": false, "street": true, "valida
v2WhjAB3PIBA8J8VxG3wEg	WiFi	no
v2WhjAB3PIBA8J8VxG3wEg	OutdoorSeating	0
v2WhjAB3PIBA8J8VxG3wEg	BusinessAcceptsCreditCards	1
v2WhjAB3PIBA8J8VxG3wEg	RestaurantsPriceRange2	2
v2WhjAB3PIBA8J8VxG3wEg	BikeParking	1
v2WhjAB3PIBA8J8VxG3wEg	WheelchairAccessible	1
duHFB87uNSXImQmvBh87Q	RestaurantsTableService	0

(Output limit exceeded, 25 of 10000 total rows shown)

ii. Business table = 10000

1	Select *
2	From Business
3	

Run

Reset

id	name	neighborhood	ad
--6MefnULPED_I942VcFNA	John's Chinese BBQ Restaurant		32
--7zmmkVg-IMGaXbuVd0SQ	Primal Brewery		16
--8LPVSo5i00o61X01sV9A	Valley Bone and Joint Specialists		39
--9e10NYQuAa-CB_Rrw7Tw	Delmonico Steakhouse	The Strip	33
--9QQlMTbFzLJ_oT-ON3Xw	Great Clips		18
--ab39IjZR_xUf81WlTyHg	Famous Footwear		18
--cgVkbwTiga3OYTkymKqA	Eazor's Auto Salon		61
--cjBEbXMI2obtaRHNSFrA	Howl at the Moon	Downtown	12
--cZ6Hhc9F7VvkKxHmVZSQ	Pio Pio	Dilworth	14
--DaPTJW3-tB1vP-PfdTEg	Sunnyside Grill	Corso Italia	12
--DdmeR16TRb3LsjG0ejrQ	World Food Championships	The Strip	36
--e8PjCNhEz32pprnPhCwQ	Lucky's Pet Grooming & Boutique	Downtown	14
--EX4rRznJrltyn-34Jz1w	Bath & Body Works		68
--FBCX-N37CMYDfs7908nw	The Bar At Bermuda & St. Rose	Southeast	11
--FLdgM0GNpXVMn74ppCGw	Welch Physical Therapy		30
--g-a85VwrdZJNf0R95GcQ	Kabab House		35
--GM_ORV2cYS-h38DSaCLw	Mm Mm Pizza		40
--i1tTcgg8i4cPkd-h5hDg	Lake Erie Nature & Science Center		28
--I7YYLada0tSLkORTHb5Q	Happy Moose Bar and Grill		94
--j-kaNMCo1-DYzddCsA5Q	Q's Nails	Eastland	30
--KCl2FvVQpvjzmZSPyviA	Hungry Howie's Pizza		64
--kinfHwmtDjz03g8B8z8Q	The Manor - Boutique Salon	Mount Pleasant and Davisville	65
--KQsXc-clk07oHRqGzSzg	Sam's Club		15
--lpHmVmkCuji0ZrpHTXEA	SkinRN	Anthem	26
--LY7PrnEegg187vnPCjQw	International Newsagents	Old Town	35

(Output limit exceeded, 25 of 10000 total rows shown)

iii. Category table = 10000

```

1  Select *
2  From Category
3

```

business_id	category
YDf95gJZaq05wvo7hTQbbQ	Shopping
YDf95gJZaq05wvo7hTQbbQ	Shopping Centers
mLwM-h2YhXl2NCgdS84_Bw	Food
mLwM-h2YhXl2NCgdS84_Bw	Soul Food
mLwM-h2YhXl2NCgdS84_Bw	Convenience Stores
mLwM-h2YhXl2NCgdS84_Bw	Restaurants
v2WhjAB3PIBA8J8VxG3wEg	Food
v2WhjAB3PIBA8J8VxG3wEg	Coffee & Tea
CVtCbSB1zUcUWg-9TNGTuQ	Professional Services
CVtCbSB1zUcUWg-9TNGTuQ	Matchmakers
duHFB87uNSXImQmvBh87Q	Sandwiches
duHFB87uNSXImQmvBh87Q	Restaurants
uUEMrhJiL1a1pCA_I1SU7Q	Shopping
uUEMrhJiL1a1pCA_I1SU7Q	Tobacco Shops
2eJEUJIP54tex7T9YOcLSw	Chiropractors
2eJEUJIP54tex7T9YOcLSw	Health & Medical
fEylCY3UEH8YJ0Xa7lu61A	Automotive
fEylCY3UEH8YJ0Xa7lu61A	Oil Change Stations
fEylCY3UEH8YJ0Xa7lu61A	Car Wash
fEylCY3UEH8YJ0Xa7lu61A	Auto Detailing
kFtuYklkAIImYw8RZAieGw	Jewelry Repair
kFtuYklkAIImYw8RZAieGw	Gold Buyers
kFtuYklkAIImYw8RZAieGw	Local Services
kFtuYklkAIImYw8RZAieGw	Shopping
kFtuYklkAIImYw8RZAieGw	Appraisal Services

(Output limit exceeded, 25 of 10000 total rows shown)

iv. Checkin table = 10000

```

1  Select *
2  From Checkin
3

```

business_id	date	count
7KPBkxAOEt3QeIL9PEErg	Thursday-21:00	4
7KPBkxAOEt3QeIL9PEErg	Thursday-1:00	1
7KPBkxAOEt3QeIL9PEErg	Thursday-4:00	1
7KPBkxAOEt3QeIL9PEErg	Thursday-2:00	1
7KPBkxAOEt3QeIL9PEErg	Thursday-20:00	2
7KPBkxAOEt3QeIL9PEErg	Thursday-22:00	1
7KPBkxAOEt3QeIL9PEErg	Thursday-19:00	1
7KPBkxAOEt3QeIL9PEErg	Thursday-15:00	2
7KPBkxAOEt3QeIL9PEErg	Thursday-13:00	1
7KPBkxAOEt3QeIL9PEErg	Thursday-23:00	2
7KPBkxAOEt3QeIL9PEErg	Wednesday-11:00	2
7KPBkxAOEt3QeIL9PEErg	Wednesday-13:00	2
7KPBkxAOEt3QeIL9PEErg	Wednesday-14:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-17:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-6:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-2:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-0:00	2
7KPBkxAOEt3QeIL9PEErg	Wednesday-1:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-21:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-18:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-19:00	1
7KPBkxAOEt3QeIL9PEErg	Wednesday-20:00	2
7KPBkxAOEt3QeIL9PEErg	Sunday-18:00	1
7KPBkxAOEt3QeIL9PEErg	Sunday-16:00	1
7KPBkxAOEt3QeIL9PEErg	Sunday-14:00	1

(Output limit exceeded, 25 of 10000 total rows shown)

v. elite_years table = 10000

```

1  Select *
2  From elite_years
3

```

user_id	year
lsSiIjAKVl-QRxEjRErBeg	2014
lsSiIjAKVl-QRxEjRErBeg	2016
lsSiIjAKVl-QRxEjRErBeg	2013
lsSiIjAKVl-QRxEjRErBeg	2011
lsSiIjAKVl-QRxEjRErBeg	2012
lsSiIjAKVl-QRxEjRErBeg	2015
lsSiIjAKVl-QRxEjRErBeg	2010
lsSiIjAKVl-QRxEjRErBeg	2017
om5ZiponkpRqUNa3pVPiRg	2014
om5ZiponkpRqUNa3pVPiRg	2017
om5ZiponkpRqUNa3pVPiRg	2011
om5ZiponkpRqUNa3pVPiRg	2012
om5ZiponkpRqUNa3pVPiRg	2015
om5ZiponkpRqUNa3pVPiRg	2009
om5ZiponkpRqUNa3pVPiRg	2013
om5ZiponkpRqUNa3pVPiRg	2007
om5ZiponkpRqUNa3pVPiRg	2016
om5ZiponkpRqUNa3pVPiRg	2006
om5ZiponkpRqUNa3pVPiRg	2010
om5ZiponkpRqUNa3pVPiRg	2008
D-ydMTPGwXTVm4_jjp0k9g	2016
D-ydMTPGwXTVm4_jjp0k9g	2017
PcvbBOC0cs6_suRDH7TSTg	2017
PcvbBOC0cs6_suRDH7TSTg	2016
N2arP_u4sMGLgOZhA6ZFoQ	2017

(Output limit exceeded, 25 of 10000 total rows shown)

vi. friend table = 10000

```

1  Select *
2  From friend
3

```

user_id	friend_id
lsSiIjAKVl-QRxBKjRErBeg	M19NwFwAXKRZzt8koF11hQ
lsSiIjAKVl-QRxBKjRErBeg	QRcMZ8pJJBBZaKubHOoMDQ
lsSiIjAKVl-QRxBKjRErBeg	uimsjcHoBnXz1MAKGvB26w
lsSiIjAKVl-QRxBKjRErBeg	v325XGF-19da74ZMwEjyoA
lsSiIjAKVl-QRxBKjRErBeg	vP5ajc1oGURsNvCXewsnDw
lsSiIjAKVl-QRxBKjRErBeg	9nSutZ0liE9Vg4XVGEx1HA
lsSiIjAKVl-QRxBKjRErBeg	--2vR0DIsmQ6WfcSzKWigw
lsSiIjAKVl-QRxBKjRErBeg	LDJ51sk5SJXovRI2yQZimA
lsSiIjAKVl-QRxBKjRErBeg	3R_dB9VQ_D3WPJEW7pmorA
lsSiIjAKVl-QRxBKjRErBeg	8drMKNHwavs2g6uf0pLtvG
lsSiIjAKVl-QRxBKjRErBeg	wOGfOjBaP-1CS1NW_En2LQ
lsSiIjAKVl-QRxBKjRErBeg	AK2-Pvb6E9vgeXWlyY4Jxog
lsSiIjAKVl-QRxBKjRErBeg	DbUSCSMQwD3eiAre0Ueu8A
lsSiIjAKVl-QRxBKjRErBeg	B_2qev6exPELs7Zn04iljg
lsSiIjAKVl-QRxBKjRErBeg	LQALTuDeCRLwR9N0xUW55A
lsSiIjAKVl-QRxBKjRErBeg	kSUU18CH2BRPLK1uUz1Wg
lsSiIjAKVl-QRxBKjRErBeg	M-HINGCHOnaQkKq8WdTRMA
lsSiIjAKVl-QRxBKjRErBeg	9lWyDOySHcc6Jiqp2-EPuW
lsSiIjAKVl-QRxBKjRErBeg	j2Eu9pE22Rp_DRoSp3KgQg
lsSiIjAKVl-QRxBKjRErBeg	neuz9oCcHiw4k-jltcC1BA
lsSiIjAKVl-QRxBKjRErBeg	PRQxRp1IFHPBlbXeDwG3mA
lsSiIjAKVl-QRxBKjRErBeg	t9vCxltuXJ941V8pplWVsVQ
lsSiIjAKVl-QRxBKjRErBeg	pYK8JuBylomxLIwwyv0Iyw
lsSiIjAKVl-QRxBKjRErBeg	WTLPH3jIWOUTFMpD4o_7Vg
lsSiIjAKVl-QRxBKjRErBeg	qAE5pJYa75gRbpC7bgI30w

(Output limit exceeded, 25 of 10000 total rows shown)

vii. hours table = 10000

```

1  Select *
2  From hours
3

```

hours	business_id
Monday 10:00-21:00	YDf95gJZaq05wvo7hTQbbQ
Tuesday 10:00-21:00	YDf95gJZaq05wvo7hTQbbQ
Friday 10:00-21:00	YDf95gJZaq05wvo7hTQbbQ
Wednesday 10:00-21:00	YDf95gJZaq05wvo7hTQbbQ
Thursday 10:00-21:00	YDf95gJZaq05wvo7hTQbbQ
Sunday 11:00-18:00	YDf95gJZaq05wvo7hTQbbQ
Saturday 10:00-21:00	YDf95gJZaq05wvo7hTQbbQ
Monday 10:00-22:00	mLwM-h2YhX12NCgdS84_Bw
Tuesday 10:00-22:00	mLwM-h2YhX12NCgdS84_Bw
Friday 10:00-22:00	mLwM-h2YhX12NCgdS84_Bw
Wednesday 10:00-22:00	mLwM-h2YhX12NCgdS84_Bw
Thursday 10:00-22:00	mLwM-h2YhX12NCgdS84_Bw
Sunday 10:00-22:00	mLwM-h2YhX12NCgdS84_Bw
Saturday 10:00-22:00	mLwM-h2YhX12NCgdS84_Bw
Monday 10:00-19:00	v2WhjAB3PIBA8J8VxG3wEg
Tuesday 10:00-19:00	v2WhjAB3PIBA8J8VxG3wEg
Friday 10:00-19:00	v2WhjAB3PIBA8J8VxG3wEg
Wednesday 10:00-19:00	v2WhjAB3PIBA8J8VxG3wEg
Thursday 10:00-19:00	v2WhjAB3PIBA8J8VxG3wEg
Sunday 12:00-17:00	v2WhjAB3PIBA8J8VxG3wEg
Saturday 10:00-18:00	v2WhjAB3PIBA8J8VxG3wEg
Friday 9:00-17:00	CVtCbSB1zUcUWg-9TNGTuQ
Tuesday 9:00-17:00	CVtCbSB1zUcUWg-9TNGTuQ
Thursday 9:00-17:00	CVtCbSB1zUcUWg-9TNGTuQ
Wednesday 9:00-17:00	CVtCbSB1zUcUWg-9TNGTuQ

(Output limit exceeded, 25 of 10000 total rows shown)

viii. photo table = 10000

```

1  Select *
2  From photo
3

```

Run

Reset

id	business_id	caption
--0uqWanwN310kuuwJ1zjQ	XIg92ukZJn_1aiNx00musQ	Le Gadboi: Poutine with pogos and onion rings
--3gCyZ0Eo3r07tAvUbhg	VeIL_tgw7ds1-7IcnOsh0g	Their OMAKASE is AMAZING!!!! Very recommende
--3vR19cePIkGQBgCLsQkw	If6Bku2jkgPiikR6HBu-XQ	Big steak omelette
--7InqXrxwFUY6-ePZILpQ	QJatAcxYgK12p9BRZMAx7g	
--9fNU-8m06bbX03jIha_w	ICdoT0DBaprN0UReete9VQ	Chocolate Cake 08/12/17 @endoedibles on Insta
--a8uNdcCabbj7HuhX9bVQ	C9xw2AkDMtWMQ3sIDo98aA	Garlic Sauce with Pork
--Ae-P_qjpFDgaQHj2nauw	eKznX8VTfcQrjCqXpeobiw	
--Be9sfiHG00D6eWn3jcpq	X3UyZfp1YAXdS3pWpaxUIA	From website. Desperately want this...now!
--cBCIMbf6bu7_1t17PzgQ	DQXz0VTP6n14gKrop8xVvw	Inside mid afternoon.
--CTOdFZ8W5_5n8LePlIpg	24UDZTAMDUaugpkchFk60w	This was the winner--fish and veggies were ex
--cVxmD-V1T85dw6MjSQew	tzEMKoLQY-ZEE6r0FO_OgA	Breakfast on the patio at the Henhouse Cafe w
--daSIW0JaP8NaJIC0-p8A	sh3UsoLkjk01u0HlQ9_0Q	Enchiladas poblanos.
--DpaHUw76HtjHogXfLXnA	SJU-jRAZS0cXo8GUjX5Gug	Crab fried rice
--F3tlcMVkMs9Bf7auWL0g	xjBvbDmbsA0VX-vA55SbAA	Open bar and dining area with comfy seating
--Fnqhgy36ck7684WwNvFQ	gMUAN6xcuE-TbY1seFw_lw	
--FRLZTnHpdIsUGTTpv_A	JXiOluj81i7aZvkN39xpzQ	Main Dining Area
--fUE1CMj62U1Dc6S8ugjg	82hYL5WPQpvxg8ekDsB83A	
--GnwI6w4f2EhVqH3V7iDQ	Wxxvi3LZbHNIDwJ-ZimtnA	Lobby.
--GxTabLHDiUMpwUntf03A	cSSgeQQ0z2modfT7zTHJHQ	Storefront - Courtesy of Foodies Inked.
--HntFmdNPowYj-xH-s_0Q	p3YqOYELqXtLyHz9T49p_w	Lunch with the Uclaray Mafia. Party of 30 ple
--hsN4HeljWID1UhwNNFA	Gu8D1YoKEnqVNNX00qkEaw	
--ifyOhCW51WtECbrsEbba	AKBSPjk_H_w8RCqCE_vUUA	Shrimp and avocado salad !
--iJGpyJEg-steFWLzJUpg	PFAvETr4Vf6UY548TwhFA	
--IMSjV2p_zmX93Dah-xXw	owxcYAUca_6K6BAUX4V48w	
--je29Go4V-WYQw0TvtypA	BjrKNWhtQkedHw8hP_0Bjg	Salmon skin hand roll

(Output limit exceeded, 25 of 10000 total rows shown)

ix. review table = 10000

--007YIsRSNb33JY0dyqpg	4	2011-08-29 00:00:00	The movie screen is huge, so no matter w The Regal 18 movie theater might become Super good food! My friends and I ordered 2 lbs of shrimp We ate it alllllll up. My only concern was that the crabs were But other than the difficult cracking, t And it was not primetime when we went so Hope they're on top of their game when i Wow i don't get all the bad reviews.
--01ogTXqLH2TzILZfrEYw	5	2014-08-04 00:00:00	I came with 2 friends and had the most a I simply had 2 tiger rolls and breaded c We had an awesome server (Joey) and we w Price for tiger roll i believe was \$24 a Allegiant is a disaster. Their fares ar The issue is not that there was a proble
--03fUVGimHb46r8XKjn2A	1	2017-06-27 00:00:00	Update: 6:30 pm and the boarding estima I went here twice and after leaving with My first visit I specifically requested As another reviewer noted, they tend to Only one lady at the front desk was very The last in the chain of events to make Nothing this office did was horrible, bu
--04HOM-bCCdrqOsnlH57w	2	2016-02-04 00:00:00	

(Output limit exceeded, 25 of 10000 total rows shown)

x. tip table = 10000

1	Select *
2	From tip
3	

Run
Reset

user_id	business_id	text
zcTZk7OG8ovAmh_fenH21g	tJRD1l5yqpZwehenzE2cSg	Get here early enough to have dinner.
ZcLKXikTHYOnYt5VYR05sg	jH19V2I9fIs1nNhDzPmdKA	Great breakfast large portions and friendly w
oayHjq8Bbh18Zhu0bpyzSuw	dAa0h82yrnHzVmsCkN4YvQ	Nice place. Great staff. A fixture in the to
ulQ8Nyj7jCUR8M83SUMoRQ	dAa0h82yrnHzVmsCkN4YvQ	Happy hour 5-7 Monday - Friday
ulQ8Nyj7jCUR8M83SUMoRQ	ESz03Av0b1_TzK0iqzbQYQ	Parking is a premium, keep circling, you will
ulQ8Nyj7jCUR8M83SUMoRQ	k7WRPbdD7rztjHcGGKejlw	Homemade pasta is the best in the area
ulQ8Nyj7jCUR8M83SUMoRQ	k7WRPbdD7rztjHcGGKejlw	Excellent service, staff is dressed professio
ulQ8Nyj7jCUR8M83SUMoRQ	Sqw3igh1_Png336Vib5DUA	Come early on Sunday's to avoid the rush
ulQ8Nyj7jCUR8M83SUMoRQ	KNpcPGqDORDdvtekXd348w	Love their soup!
ulQ8Nyj7jCUR8M83SUMoRQ	KNpcPGqDORDdvtekXd348w	Soups are fantastic!
ulQ8Nyj7jCUR8M83SUMoRQ	KNpcPGqDORDdvtekXd348w	Thursday night is \$5 burger night
ulQ8Nyj7jCUR8M83SUMoRQ	8qNOI6Q1-rJrVWMD58tz6w	Very good for lunch, bread and soup are excel
ulQ8Nyj7jCUR8M83SUMoRQ	8qNOI6Q1-rJrVWMD58tz6w	Tuesday spaghetti special all day \$5.59 incl
ulQ8Nyj7jCUR8M83SUMoRQ	--ujyvoQlwVoBgMYtADiLA	Sunday \$.55 bone-in wings
ulQ8Nyj7jCUR8M83SUMoRQ	ulQ8Nyj7jCUR8M83SUMoRQ	Monday \$.55 boneless wings
ulQ8Nyj7jCUR8M83SUMoRQ	FCUVjQf762no86Uzcbv1Tg	There isn't a bad seat in the house!
ulQ8Nyj7jCUR8M83SUMoRQ	Tl_jT2a0bRMac5_YW65GPg	Nice happy hour m-f 5-7pm and th-fr 10pm-12am
ulQ8Nyj7jCUR8M83SUMoRQ	Tl_jT2a0bRMac5_YW65GPg	Always have something new on the inter menu,
ulQ8Nyj7jCUR8M83SUMoRQ	C5NHHW0sNm7eaQBwvQJkkw	Brisket sandwich not bad but ask for extra na
ulQ8Nyj7jCUR8M83SUMoRQ	zJwm3DThV4WUikY/Nv68Ahw	Double wide does offer free charging stations
FHGJ3xnxKYoSZqmQJr_pLg	dAa0h82yrnHzVmsCkN4YvQ	Very good, their side salads are amazing, ver
FHGJ3xnxKYoSZqmQJr_pLg	7B-xM_3t9NKeBmkEodw3CQ	Good everything of course
FHGJ3xnxKYoSZqmQJr_pLg	k7WRPbdD7rztjHcGGKejlw	The best food ever!!! Homemade everything jus
FHGJ3xnxKYoSZqmQJr_pLg	Zt2cMfngRaSbYdc0Sh4jA	Always sooooo good!!
TvkqJ8YEIsTb16RnnrNyfq	JzB7NITHQ7gVHGvZ1ntgIQ	Black Angus and the Roast beef :)
TvkqJ8YEIsTb16RnnrNyfq	h14GmwZ8rXum9fXF__wt3w	Expensive, but convenient for hotel stays

(Output limit exceeded, 25 of 10000 total rows shown)

xi. user table = 10000

1	Select *
2	From user
3	

Run

Reset

id	name	review_count	yelping_since	useful	funny	co
---1lKK3aK0uomHnwAkAow	Monera	245	2007-06-04 00:00:00	67	22	
---94vtJ_5o_nikEs6hUjg	Joe	2	2016-05-27 00:00:00	0	0	
---cu1hq55BP9DwVXXKHZg	Jeb	57	2009-04-18 00:00:00	34	14	
---fhiwiwBYrvqhpXgcwDQ	Jed	8	2011-04-20 00:00:00	2	3	
---PLwSf5gKdIoVnyRHgBA	Rae	2	2015-07-31 00:00:00	1	0	
---udAKDsnyQXmzbwQNSw	Carolyn	43	2014-07-12 00:00:00	1	0	
---0kuuLmuYBe3Rmu0Iycww	Talia	26	2010-03-08 00:00:00	10	2	
---0RtXvcOIE4XbErYca6Rw	Ryan	2	2013-05-30 00:00:00	0	0	
---0sXNBv6IizZXuV-nl0Aw	Joe	1	2013-01-09 00:00:00	0	0	
---0WZ5gklofbUIodJukfaQ	Scott	7	2013-02-19 00:00:00	0	0	
---104qdWvE99vaoIsj9ZJQ	John	3	2016-04-26 00:00:00	0	0	
---1av6NdbEbMiuBr7Aup9A	Ron	9	2010-09-26 00:00:00	0	0	
---1mPJZdSY9KluaBYAGboQ	Bryan	5	2011-07-04 00:00:00	0	0	
---26jc8nC3By4-7r3ZtmiQ	Patti	2	2014-08-03 00:00:00	15	13	
---2bpE5vyR-2hAP7sZZ4lA	Gary	23	2015-10-12 00:00:00	0	0	
---2HUmLkcNHZp0Xw6AMBpG	Kristin	28	2016-07-28 00:00:00	7	1	
---2vR0DismQ6WfcSzKWigw	Harald	1153	2012-11-27 00:00:00	122921	122419	1228
---3B8LdT1NCD-bPkwS5-5g	Cynthia	4	2016-11-10 00:00:00	0	0	
---3l8wysfp49Z2TLnyT0vg	Benjamin	111	2013-12-14 00:00:00	97	57	
---3oMd6gjXpAzhjLBRsVCQ	Mnme	2	2010-03-22 00:00:00	1	0	
---3WAS23LcIXtxyFULJHTA	Kristie	213	2010-05-02 00:00:00	63	6	
---4lc9Tl0C9OGewIR7Qyzg	Tamaki	239	2011-07-03 00:00:00	64	15	
---44NndtngXMzsyN7ju6Q	Austin	2	2013-01-22 00:00:00	0	0	
---4q8EyqThydQm-eKZpS-A	Kiristen	400	2008-01-07 00:00:00	405	313	
---4rAAfZnEIAKJE80aIiyg	Mesut	25	2013-09-14 00:00:00	12	5	

(Output limit exceeded, 25 of 10000 total rows shown)

2. Find the total distinct records by either the foreign key or primary key for each table. If two foreign keys are listed in the table, please specify which foreign key. Note: Primary Keys are denoted in the ER-Diagram with a yellow key icon.

- i. Business = Primary key: id; total distinct 10000 records.

1	Select COUNT (DISTINCT Id)
2	From Business
3	

COUNT (DISTINCT Id)
10000

- ii. Hours = Foreign key: business_id; total distinct 1562 records.

```
1  Select COUNT (DISTINCT business_id)
2  From Hours
3
```

COUNT (DISTINCT business_id)
1562

- iii. Category = Foreign key: business_id; total distinct 2643 records.

```
1  Select COUNT (DISTINCT business_id)
2  From Category
3
```

COUNT (DISTINCT business_id)
2643

- iv. Attribute = Foreign key: business_id; total distinct 1115 records.

```
1  Select COUNT (DISTINCT business_id)
2  From Attribute
3
```

COUNT (DISTINCT business_id)
1115

- v. Review = Foreign key: business_id; total distinct 8090 records.

1	Select COUNT (DISTINCT business_id)
2	From Review
3	

+	-----+
	COUNT (DISTINCT business_id)
+	-----+
	8090
+	-----+

vi. Checkin = Foreign key: business_id; total distinct 493 records.

1	Select COUNT (DISTINCT business_id)
2	FROM Checkin

+	-----+
	COUNT (DISTINCT business_id)
+	-----+
	493
+	-----+

vii. Photo = Foreign key: business_id; total distinct 6493 records.

1	Select COUNT (DISTINCT business_id)
2	FROM Photo

+	-----+
	COUNT (DISTINCT business_id)
+	-----+
	6493
+	-----+

viii. Tip = Foreign key: business_id; total distinct 3979 records.

1	Select COUNT (DISTINCT business_id)
2	FROM Tip

+	-----+
	COUNT (DISTINCT business_id)
+	-----+
	3979
+	-----+

ix. User = Foreign key: business_id; total distinct 10000 records.

1	Select COUNT (DISTINCT id)
2	FROM User

+	-----	+
	COUNT (DISTINCT id)	
+	-----	+
	10000	
+	-----	+

x. Friend = Foreign key: user_id; total distinct 11 records.

1	SELECT COUNT (DISTINCT user_id)
2	FROM Friend

+	-----	+
	COUNT (DISTINCT user_id)	
+	-----	+
	11	
+	-----	+

xii. Elite_years = Foreign key: user_id; total distinct 2780 records.

1	SELECT COUNT (DISTINCT user_id)
2	FROM Elite_years

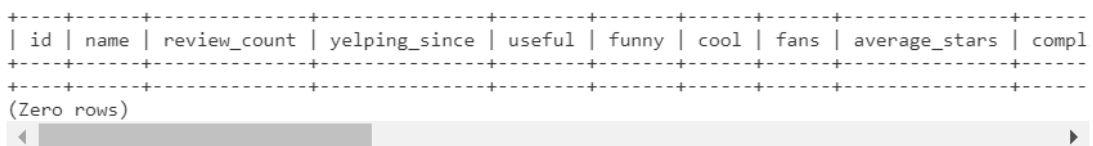
+	-----	+
	COUNT (DISTINCT user_id)	
+	-----	+
	2780	
+	-----	+

3. Are there any columns with null values in the Users table? Indicate "yes," or "no."

Answer: No

SQL code used to arrive at answer:

```
SELECT *
FROM User
WHERE User.id IS NULL OR
User.name IS NULL OR
User.review_count IS NULL OR
User.yelping_since IS NULL OR
User.useful IS NULL OR
User.funny IS NULL OR
User.cool IS NULL OR
User.fans IS NULL OR
User.average_stars IS NULL OR
User.compliment_hot IS NULL OR
User.compliment_more IS NULL OR
User.compliment_profile IS NULL OR
User.compliment_cute IS NULL OR
User.compliment_list IS NULL OR
User.compliment_note IS NULL OR
User.compliment_plain IS NULL OR
User.compliment_cool IS NULL OR
User.compliment_funny IS NULL OR
User.compliment_writer IS NULL OR
User.compliment_photos IS NULL
```



```
+---+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| id | name | review_count | yelping_since | useful | funny | cool | fans | average_stars | compl |
+---+-----+-----+-----+-----+-----+-----+-----+-----+-----+
(Zero rows)
```

4. For each table and column listed below, display the smallest (minimum), largest (maximum), and average (mean) value for the following fields:

i. Table: Review, Column: Stars

min: 1 max: 5 avg: 3.7082

```
1 SELECT MIN (Stars)
2 FROM Review
```

```
+-----+
| MIN (Stars) |
+-----+
|          1 |
+-----+
```

```
1 SELECT MAX (Stars)
2 FROM Review
```

```
+-----+
| MAX (Stars) |
+-----+
|          5 |
+-----+
```

```
1 SELECT AVG (Stars)
2 FROM Review
```

```
+-----+
| AVG (Stars) |
+-----+
|    3.7082 |
+-----+
```

ii. Table: Business, Column: Stars

min: 1 max: 5 avg: 3.6549

```

1  SELECT MIN (Stars)
2  FROM Business

```

```

+-----+
| MIN (Stars) |
+-----+
|          1.0 |
+-----+

```

```

1  SELECT MAX (Stars)
2  FROM Business

```

```

+-----+
| MAX (Stars) |
+-----+
|          5.0 |
+-----+

```

```

1  SELECT AVG (Stars)
2  FROM Business

```

```

+-----+
| AVG (Stars) |
+-----+
|       3.6549 |
+-----+

```

iii. Table: Tip, Column: Likes

min: 0 max: 2 avg: 0.0144


```

1  SELECT MIN (Likes)
2  FROM Tip

```

```

+-----+
| MIN (Likes) |
+-----+
|           0 |
+-----+

```

```

1  SELECT MAX (Likes)
2  FROM Tip

```

```

+-----+
| MAX (Likes) |
+-----+
|           2 |
+-----+

```

```

1  SELECT AVG (Likes)
2  FROM Tip

```

```

-----+
AVG (Likes) |
-----+
    0.0144 |
-----+

```

iv. Table: Checkin, Column: Count

min: 1 max: 53 avg: 1.9414

```

1  SELECT MIN (Count)
2  FROM Checkin

```

```

+-----+
| MIN (Count) |
+-----+
|           1 |
+-----+

```

```

1  SELECT MAX (Count)
2  FROM Checkin

```

```

+-----+
| MAX (Count) |
+-----+
|          53 |
+-----+

```

```

1  SELECT AVG (Count)
2  FROM Checkin

```

```

+-----+
| AVG (Count) |
+-----+
|       1.9414 |
+-----+

```

v. Table: User, Column: Review_count

min: 0 max: 2000 avg: 24.2995

```

1  SELECT MIN (Review_count)
2  FROM User

```

```

+-----+
| MIN (Review_count) |
+-----+
|                   0 |
+-----+

```

```

1  SELECT MAX (Review_count)
2  FROM User

```

```

+-----+
| MAX (Review_count) |
+-----+
|                 2000 |
+-----+

```

```

1  SELECT AVG (Review_count)
2  FROM User

```

```

+-----+
| AVG (Review_count) |
+-----+
|                24.2995 |
+-----+

```

5. List the cities with the most reviews in descending order:

SQL code used to arrive at answer:

```
SELECT business.review_count, business.city
FROM business
ORDER BY business.review_count DESC
```

Copy and Paste the Result Below:

review_count	city
3873	Las Vegas
1757	Montréal
1549	Gilbert
1410	Las Vegas
1389	Las Vegas
1252	Las Vegas
1116	Las Vegas
1084	Las Vegas
961	Las Vegas
902	Gilbert
864	Las Vegas
823	Scottsdale
821	Las Vegas
786	Las Vegas
785	Henderson
778	Toronto
768	Las Vegas
758	Las Vegas
726	Scottsdale
723	Cleveland
720	Las Vegas
715	Charlotte
711	Phoenix
706	Las Vegas
700	Phoenix

(Output limit exceeded, 25 of 10000 total rows shown)

6. Find the distribution of star ratings to the business in the following cities:

i. Avon

SQL code used to arrive at answer:

```
SELECT review.stars, COUNT(*) as stars_count, 'Avon'
FROM business LEFT JOIN review
WHERE business.city = 'Avon'
GROUP BY review.stars
ORDER BY review.stars DESC
```

Copy and Paste the Resulting Table Below (2 columns – star rating and count):

stars	stars_count	'Avon'
5	41510	Avon
4	24010	Avon
3	12190	Avon
2	8370	Avon
1	13920	Avon

ii. Beachwood

SQL code used arrive at answer:

```
SELECT review.stars, COUNT(*) as stars_count, 'Beachwood'
FROM business LEFT JOIN review
WHERE business.city = 'Beachwood'
GROUP BY review.stars
ORDER BY review.stars DESC
```

Copy and Paste the Resulting Table Below (2 columns – star rating and count):

stars	stars_count	'Beachwood'
5	58114	Beachwood
4	33614	Beachwood
3	17066	Beachwood
2	11718	Beachwood
1	19488	Beachwood

7. Find the top 3 users based on their total number of reviews:

SQL code used to arrive at answer:

```
SELECT user.name, user.review_count
FROM user
ORDER BY user.review_count DESC
LIMIT 3
```

Copy and Paste the Result Below:

name	review_count
Gerald	2000
Sara	1629
Yuri	1339

8. Does posing more reviews correlate with more fans? Please explain your findings and interpretation of the results:

To have a minimum number of fans, people need to set reviews. However, the top users with the review don't have more fans than those with a medium number of reviews. For example, the user "Mimi" has the highest number of fans (497) and 968 reviews. It's about one thousand reviews less than "Gerald," which is on the top regarding the number of reviews but only has 253 fans.

As can be seen from the following screen shots:

name	review_count	fans
Sonnenschein1	0	0
svenher	0	0
Schweinefe	0	0
Luke	0	0
Limon-Du	0	0
ab	0	0
torstenbec	0	0
snek	0	0
Joe	1	0
Lyndsey	1	0
Qi	1	0
Sachin	1	0
Mary	1	0
Gwen	1	0
Kimmie	1	0
Rosa Maria	1	0
Carl	1	0
Michael	1	0
Tony	1	0
Regina	1	0
A	1	0
Sa	1	0
Mishelle	1	0
Rachel	1	0

name	review_count	fans
Gerald	2000	253
Sara	1629	50
Yuri	1339	76
.Hon	1246	101
William	1215	126
Harald	1153	311
eric	1116	16
Roanna	1039	104
Mimi	968	497
Christine	930	173
Ed	904	38
Nicole	864	43
Fran	862	124
Mark	861	115
Christina	842	85
Dominic	836	37
Lissa	834	120
Lisa	813	159
Alison	775	61
Sui	754	78
Tim	702	35
L	696	10
Angela	694	101
Crissy	676	25
Lyn	675	45

9. Are there more reviews with the word "love" or with the word "hate" in them?

Answer: more the word "love".

SQL code used to arrive at answer:

```
SELECT review.text
FROM review
WHERE review.text LIKE '%love%'

-- Have 1780 with the words Love or love
```

```
SELECT review.text
FROM review
WHERE review.text LIKE '%hate%'

-- Have 232 with the word hate
```

10. Find the top 10 users with the most fans:

SQL code used to arrive at answer:

```
SELECT user.name, user.fans
FROM user
ORDER BY user.fans DESC
LIMIT 10;
```

Copy and Paste the Result Below:

name	fans
Amy	503
Mimi	497
Harald	311
Gerald	253
Christine	173
Lisa	159
Cat	133
William	126
Fran	124
Lissa	120

Part 2: Inferences and Analysis

1. Pick one city and category of your choice and group the businesses in that city or category by their overall star rating. Compare the businesses with 2-3 stars to the businesses with 4-5 stars and answer the following questions. Include your code.

I picked **Phoenix** here is the presentation of the categories in Phoenix:

```
SELECT category.category, COUNT(DISTINCT business.id) as business_count
FROM business
LEFT JOIN category ON business.id = category.business_id
WHERE business.city = 'Phoenix'
GROUP BY category.category
ORDER BY business_count DESC;
```

Output:

category	business_count
None	980
Restaurants	6
Food	4
Home Services	4
American (Traditional)	3
Health & Medical	3
Bars	2
Beauty & Spas	2
Burgers	2
Chiropractors	2
Contractors	2
Fast Food	2
Nightlife	2
Shopping	2
Active Life	1
American (New)	1
Apartments	1
Auto Detailing	1
Automotive	1
Barbeque	1
Breakfast & Brunch	1
Car Wash	1
Coffee & Tea	1
Convenience Stores	1
Doctors	1

(Output limit exceeded, 25 of 51 total rows shown)

Then I choose the category "restaurants" in phoenix.

```
SELECT business.name,
business.stars
FROM business
LEFT JOIN category ON business.id = category.business_id
WHERE business.city = 'Phoenix' and category = 'Restaurants'
ORDER BY business.stars DESC;
```

Output:

name	stars
Charlie D's Catfish & Chicken	4.5
Matt's Big Breakfast	4.0
Bootleggers Modern American Smokehouse	4.0
Five Guys	3.5
Gallagher's	3.0
McDonald's	2.0

- i. Do the two groups you chose to analyze have a different distribution of hours?

Yes, we can see that from 6 restaurants 2 with the 4.5 and 4 stars have 11:00-18:00 and 11:00-22:00, respectively. And each of the 3 restaurants that have lower than 4 stars have different hours.

The SQL code:

```
SELECT business.name, business.stars, hours.hours
FROM business
LEFT JOIN category ON business.id = category.business_id
LEFT JOIN hours ON business.id = hours.business_id
WHERE business.city = 'Phoenix' AND category.category = 'Restaurants'
GROUP BY business.name
ORDER BY business.stars DESC;
```

The output:

name	stars	hours
Charlie D's Catfish & Chicken	4.5	Saturday 11:00-18:00
Bootleggers Modern American Smokehouse	4.0	Saturday 11:00-22:00
Matt's Big Breakfast	4.0	None
Five Guys	3.5	Saturday 10:00-22:00
Gallagher's	3.0	Saturday 9:00-2:00
McDonald's	2.0	Saturday 5:00-0:00

ii. Do the two groups you chose to analyze have a different number of reviews?

Yes, as can be seen from the output; in general, the group of 4 stars and above have more reviews than the group of 2-3.5 stars.

The SQL code:

```
SELECT business.name, business.stars, hours.hours, business.review_count
FROM business
LEFT JOIN category ON business.id = category.business_id
LEFT JOIN hours ON business.id = hours.business_id
WHERE business.city = 'Phoenix' AND category.category = 'Restaurants'
GROUP BY business.name
ORDER BY business.stars DESC;
```

The output:

name	stars	hours	review_count
Charlie D's Catfish & Chicken	4.5	Saturday 11:00-18:00	7
Bootleggers Modern American Smokehouse	4.0	Saturday 11:00-22:00	431
Matt's Big Breakfast	4.0	None	188
Five Guys	3.5	Saturday 10:00-22:00	63
Gallagher's	3.0	Saturday 9:00-2:00	60
McDonald's	2.0	Saturday 5:00-0:00	8

iii. Are you able to infer anything from the location data provided between these two groups? Explain.

Yes, by putting those 6 restaurants on google maps. I notice that the 2 most reviews restaurants ("Bootleggers Modern American Smokehouse" with 431 reviews and "Matt's Big Breakfast" with 188 reviews) is located near to "Phoenix Mountains Preserve".

SQL code used for analysis:

```
SELECT business.name, business.stars, business.review_count,
business.address, business.latitude, business.longitude
FROM business
LEFT JOIN category ON business.id = category.business_id
LEFT JOIN hours ON business.id = hours.business_id
WHERE business.city = 'Phoenix' AND category.category = 'Restaurants'
GROUP BY business.name
ORDER BY business.stars DESC;
```

The output:

name	stars	review_count	address	address	latitude	longitude
Charlie D's Catfish & Chicken	4.5	7	1153 E Jefferson St	1153 E Jefferson St	33.4468	-112.057
Bootleggers Modern American Smokehouse	4.0	431	3375 E Shea Blvd	3375 E Shea Blvd	33.5818	-112.008
Matt's Big Breakfast	4.0	188	3118 E Camelback Rd	3118 E Camelback Rd	33.511	-112.015
Five Guys	3.5	63	2641 N 44th St, Ste 100	2641 N 44th St, Ste 100	33.478	-111.986
Gallagher's	3.0	60	751 E Union Hls Dr	751 E Union Hls Dr	33.6536	-112.064
McDonald's	2.0	8	1850 S 7th St	1850 S 7th St	33.4297	-112.066

2. Group business based on the ones that are open and the ones that are closed. What differences can you find between the ones that are still open and the ones that are closed? List at least two differences and the SQL code you used to arrive at your answer.

i. Difference 1:

By looking only at the 6 restaurants found only one closed which had only 7 reviews.

The code:

```
SELECT business.is_open, business.name, business.stars,
business.review_count,
business.address
FROM business
LEFT JOIN category ON business.id = category.business_id
LEFT JOIN hours ON business.id = hours.business_id
WHERE business.city = 'Phoenix' AND category.category = 'Restaurants'
GROUP BY business.name
```

```
ORDER BY business.stars DESC;
```

The output:

is_open	name	stars	review_count
0	Charlie D's Catfish & Chicken	4.5	7
1	Bootleggers Modern American Smokehouse	4.0	431
1	Matt's Big Breakfast	4.0	188
1	Five Guys	3.5	63
1	Gallagher's	3.0	60
1	McDonald's	2.0	8

By looking at a broader view of all businesses in Phoenix we can notice a trend that less review business is likely to be close.

The code:

```
SELECT business.is_open, business.name, business.stars,
business.review_count,
business.address
FROM business
LEFT JOIN category ON business.id = category.business_id
LEFT JOIN hours ON business.id = hours.business_id
WHERE business.city = 'Phoenix'
GROUP BY business.name
ORDER BY business.review_count ASC;
```

ii. Difference 2:

The less stars the likely the business will be closed.

SQL code used for analysis:

```
SELECT business.is_open, business.name, business.stars,
business.review_count, business.address
FROM business
LEFT JOIN category ON business.id = category.business_id
LEFT JOIN hours ON business.id = hours.business_id
LEFT JOIN checkin ON hours.business_id = checkin.business_id
WHERE business.city = 'Phoenix'
GROUP BY business.name
ORDER BY business.stars ASC
```

The output:

is_open	name	stars	review_count	address
1	Acme Lawn Care	1.0	3	
1	Alphagraphics	1.0	4	3525 N C
0	Bed Bath & Beyond	1.0	4	10403 N
1	Charter Bus Express	1.0	14	4801 E M
1	Deer Valley Counseling - Phoenix	1.0	4	2301 W D
1	First Class Moving & Storage	1.0	6	21430 N
1	Jack In the Box	1.0	8	8225 W C
1	Lofts@10 Apartments	1.0	3	2247 E V
1	Martha's Bridal And Xv	1.0	6	1616 N 1
1	Maximum Air Care	1.0	5	2990 W C
1	Mega Furniture	1.0	7	3536 W G
0	One Hour Air Conditioning	1.0	8	4040 E R
1	Phoenix Parking Control	1.0	13	N Centra
1	Protect Your Home - ADT Authorized Premier Provider	1.0	20	4620 N 1
1	Rightsource-Humana Pharmacy	1.0	18	4302 W B
0	Ticketmaster	1.0	9	8181 S 4
1	US Pest Control	1.0	3	2432 W P
1	Ward North American	1.0	4	875 S 65
1	Yakety Yak Wireless	1.0	4	1859 W G
1	Advanced Transmission & Emissions	1.5	3	2849 E B
1	Advantage Rent A Car	1.5	40	1805 E S
1	Air Canada	1.5	3	3400 E S
1	Blast Fitness Phoenix- Indian School	1.5	8	4344 W I
1	Chase Bank	1.5	3	6030 N 1
1	Desert Schools Federal Credit Union	1.5	35	3423 E B

Here is by DESC order:

is_open	name	stars	review_count	address
1	A Safe Pool of Arizona	5.0	15	10115 E Bel
1	A-Z Septic Pumping	5.0	10	1036 E Clou
1	AZ Transport	5.0	7	
1	Accu-Care Cremation & Funerals	5.0	3	4033 N 19th
1	Acme Prints	5.0	18	705 N 7th A
1	Advanced Bio Solutions	5.0	6	24 W Camelb
1	Affordable Image	5.0	4	2515 N 7th
1	Agave Chiropractic	5.0	6	3040 N 44th
1	All Smiles Dentistry	5.0	10	4130 N 108t
1	Alphonsina	5.0	10	10210 N 32n
1	Ambassador Fine Cigars	5.0	11	10810 N Tat
1	Ammo AZ	5.0	5	2040 W Deer
0	Annie Boomer Vintage	5.0	5	908 N 6th S
1	Apache Wash Trailhead	5.0	9	E Sonoran D
1	Arcadia Premium	5.0	17	5618 E Thom
1	Arizona Art Restoration	5.0	5	1318 E Shan
1	Arizona Microgreens	5.0	4	3146 E Wier
1	Arizona Motorcycle Towing & Storage	5.0	9	2602 W Town
1	Arizona's Best Windows	5.0	26	
1	Az Rapid Motor Vehicle Services & Permit Company	5.0	3	536 E Dunla
1	Back-Health Chiropractic	5.0	19	4425 N 24th
1	Barefoot Pools Pool Service & Repair	5.0	83	1241 E Chan
1	Batter's Box Baseball Cards & Stuff	5.0	8	2855 W Cact
0	Battleground Sports	5.0	3	1745 W Deer
1	Bennett Parker Law	5.0	13	1601 N 7th

(Output limit exceeded, 25 of 945 total rows shown)

3. For this last part of your analysis, you are going to choose the type of analysis you want to conduct on the Yelp dataset and are going to prepare the data for analysis.

Ideas for analysis include: Parsing out keywords and business attributes for sentiment analysis, clustering businesses to find commonalities or anomalies between them, predicting the overall star rating for a business, predicting the number of fans a user will have, and so on. These are just a few examples to get you started, so feel free to be creative and come up with your own problem you want to solve. Provide answers, in-line, to all of the following:

i. Indicate the type of analysis you chose to do:

As the Yelp dataset includes information on businesses across various industries, including restaurants, retail stores, entertainment venues, and more. I want to analyze and understand which category is the riskiest for businesses. I will perform a basic risk assessment for the business category.

ii. Write 1-2 brief paragraphs on the type of data you will need for your analysis and why you chose that data:

To analyze and get an insight about the riskiest category, I will need a first get information about the activity of the business - if it is still active or permanently closed. As well as the category of the business it is classified to. In addition, I will need to compare what is the category that has the most active business and what is the category with business that are permanently closed.

iii. Output of your finished dataset:

category	open_businesses_count
None	8328
Restaurants	53
Shopping	25
Food	20
Health & Medical	16
Home Services	15
Beauty & Spas	12
Nightlife	12
Bars	11
Active Life	10
Local Services	10
Automotive	9
American (Traditional)	8
Hotels & Travel	8
Arts & Entertainment	7
Burgers	7
Fast Food	7
Hair Salons	6
Sandwiches	6
Doctors	5
Mexican	5
Apartments	4
Auto Repair	4
Bakeries	4
Indian	4

(Output limit exceeded, 25 of 236 total rows shown)

We can see that the restaurants category have the most active businesses (53) this insight tells us that the “Restaurants” is the category where the most new business openings.

The question now is if it is also the category with the most closed businesses – or in other words, if it is the riskiest category to open a new business.

category	closed_businesses_count
None	1488
Restaurants	18
Nightlife	8
Bars	6
Shopping	5
American (New)	3
American (Traditional)	3
Event Planning & Services	3
Food	3
Desserts	2
Gluten-Free	2
Italian	2
Japanese	2
Local Services	2
Mexican	2
Pizza	2
Sandwiches	2
Specialty Food	2
Venues & Event Spaces	2
Bakeries	1
Beauty & Spas	1
Beer	1
Breakfast & Brunch	1
Burgers	1
Carpet Cleaning	1

(Output limit exceeded, 25 of 72 total rows shown)

By taking the most notable categories into small excel table and organized it by the percentage risk.

Category	Open	Cloesed	Total	Shutdown risk
Nightlife	12	8	20	40%
Bars	11	6	17	35%
Restaurants	53	18	71	25%
Shopping	25	5	30	17%
Food	20	3	23	13%

The calculation for the shutdown risk is made by the following calculation:

Formula:

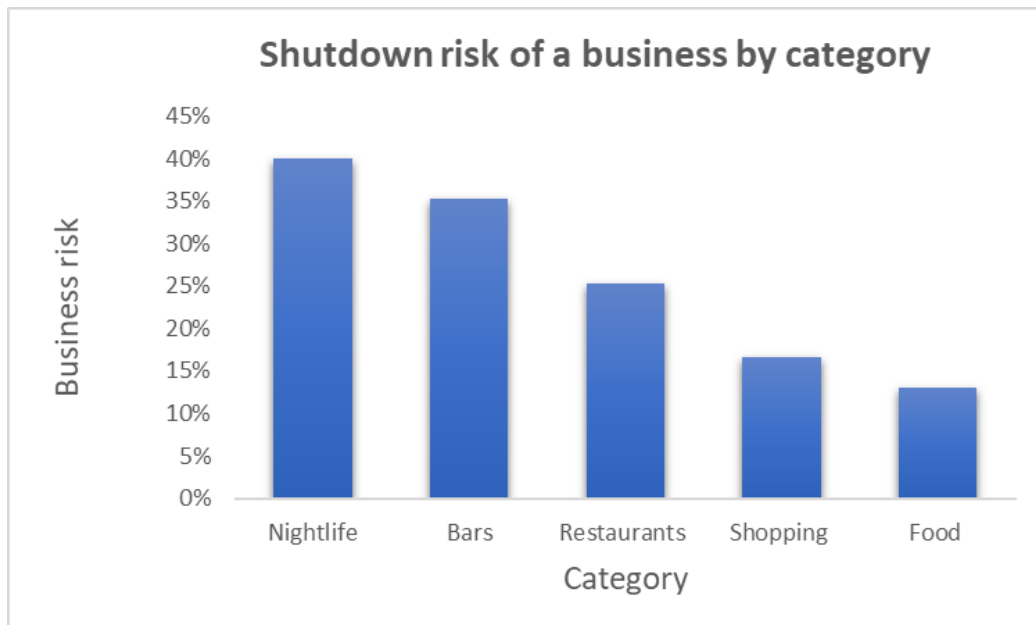
Risk of Closure = (Number of Permanently Closed Businesses in Category) / (Total Number of Businesses in Category)

Explanation:

The Risk of Closure is a measure of the probability that a business in a particular category will close down permanently. It is calculated by taking the number of businesses in that category that have already closed down permanently and dividing it by the total number of businesses in that category. This gives a percentage that represents the level of risk for

businesses in that category. A higher percentage indicates a higher risk of closure, while a lower percentage indicates a lower risk of closure.

Graph:



We can notice that the highest risk in the business category is Nightlife with 40% of closing. And the lowest of those is Food 13%. The restaurants are with 25% - means that from every 4 restaurants that are open 1 will close.

iii. Provide the SQL code you used to create your final dataset:

```
-- To get all the activley open businesses and thier category
SELECT category.category, COUNT(*) as open_businesses_count
FROM business
LEFT JOIN category ON business.id = category.business_id
WHERE business.is_open = 1
GROUP BY category.category
ORDER BY open_businesses_count DESC;

-- To get all the closed businesses and thier category
SELECT category.category, COUNT(*) as closed_businesses_count
FROM business
LEFT JOIN category ON business.id = category.business_id
WHERE business.is_open = 0
GROUP BY category.category
ORDER BY closed_businesses_count DESC;
```

Thanks in advance for the assessment and the feedback!

Alex

