Task 1.1 Extracting topics of the large dataset.

For the first task, the python toolkit was used, generating topics using LDA modeling. The dataset used was the set of all the reviews in Restaurant category, which were filtered using Java algorithms.

In order to achieve a better set of topics, the script was changed to use 10000 iterations instead of 50, because running the LDA model in the large dataset using a small number of iterations produced topics that were not useful at all. While there are still some words that are not great fits to the topics, and a few topics are not very conclusive, the execution with 10000 iterations produced some clear topics, like topic 4, topic 9 and topic 5 (the latter with some mixed words) for cuisines, and topics 8 and 10 about service and quality. The quality of the topics could have been improved if the word classification was improved to merge words with their plurals.

For presentation, it was used a customization of google charts trees, in order to make a simple, clean, yet effective presentation of the topics, where colors indicate different topics and word sizes represent the frequencies.

Image 1: Topics for All Restaurant Reviews

Task 1.2 Comparison of extracted topics.

For the second task, the python toolkit was used, generating topics using LDA modeling. Four different datasets were used. First, reviews were separated in a sample of traditional cuisines, which included Italian, Mediterranean, and French; and a sample of fast food cuisines, that included Pizza, Burger and Hot Dogs. Besides, these two sets were separated in High ratings (4 or 5 stars) and Low Ratings (1 or 2 stars).

The Model used was LDA again, with 10000 iterations in order to achieve better topics. For the topics of Traditional Cuisine reviews, the following data was obtained:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
cocktails	gelato	server	vegas	pizza
ricotta	com	opening	naan	sauce
truffle	http	pomo	masala	good
risotto	burrata	scottsdale	chicken	salad
lol	www	bar	tikka	cheese
biryani	calzone	beautiful	curry	ordered
chopped	yelp	knowledgeable	appetizers	fresh
chili	antipasto	cocktail	shawarma	delicious
filet	pros	chipotle	paneer	chicken
mushroom	hostess	federal	rice	bread

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
pita	food	pizza	view	food
hummus	service	bar	meatball	place
falafel	stars	pizzas	mushrooms	great
gyro	received	great	potatoes	good
bacon	smoked	wine	buddy	service
salad	expected	place	corn	luckily
greek	ambiance	toppings	sample	time
recommendations	exceptional	customer	pastry	friendly
chicken	cherry	las	attention	try
soft	tuesday	list	rest	really

Image 2: High Rated Reviews for Traditional Cuisines

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
steak	pizza	food	location	chicken
vegas	sauce	place	pizza	naan
ordered	salad	order	knots	masala
pork	cheese	just	place	rice
fries	good	pizza	food	curry
arugula	chicken	service	stated	meatballs
beef	ordered	minutes	drive	sauce
gravy	bread	time	pros	tikka
bitter	pasta	table	time	garlic
risotto	like	good	forever	spicy

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
shawarma	noise	cons	food	vegas
quiet	update	realize	service	pie
potential	level	sangria	place	girls
mexican	confusing	watched	good	veggie
gf	aren	delivery	bad	process
postino	fork	garbage	great	buddy
dude	bright	ready	restaurant	closing
promising	beers	coupon	try	guest
block	waitresses	industry	experience	dance
wasted	buffalo	smoke	just	shame

Image 3: Low Rating Reviews for Traditional Cuisines

It is interesting to notice that in high rating review topics, there are many different cuisines represented by the topics, and lots of the most frequent words are about specific dishes or ingredients. On the other hand, in the low rating reviews, besides having many topics related to pizza, which probably dominate the low rating reviews attributed to Italian Cuisine, the dishes and ingredients are far less specific. Many words related to dishes or ingredients are generic, like beef, pork, cheese, salad, bread, pasta, rice.

For the topics of Fast Food reviews, the following data was obtained:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
chili	pudding	steak	fries	caesar
burger	opening	finished	cheese	cocktails
milkshake	concept	cons	sauce	explained
chips	chocolate	basic	pizza	balsamic
truck	shake	smashburger	good	slaw
jalapeno	dessert	charge	burger	ricotta
egg	potatoes	ginger	ordered	helped
rare	sugar	powdered	crust	salted
patty	federal	sausages	fresh	michigan
avocado	cream	chose	salad	visits

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
service	portillo	pizza	bun	truffle
beer	http	place	ketchup	gordon
beers	com	food	strip	hostess
attentive	www	great	topping	complaint
busy	select	good	coming	fine
kitchen	yelp	burger	cheddar	actual
seated	biz_photos	like	detroit	chilli
great	calories	just	pies	fat
bar	patient	time	modern	convenient
opened	everyday	try	las	telling

Image 4: High Rating Reviews for Fast Food

Image 5: Low Rating reviews for Fast Food.

It is interesting to notice that, compared to the Traditional Cuisine, the reviews' content for fast food are rather different. The first thing to notice is that even in high rating review topics, the dominating use of generic words to refer to dishes and ingredients instead of the specific words used in traditional cuisines.

Another highlight of this second analysis is that the topics of negative reviews for fast food are not much different from the general topics for positive reviews, but there are many instances of meaningful words that describe elements not related to the food, but pricing, the size of the portions, seats, accomodations, menus and the selection. it shows that these elements are important for fast food restaurant reviewers.