What’s Cooking?

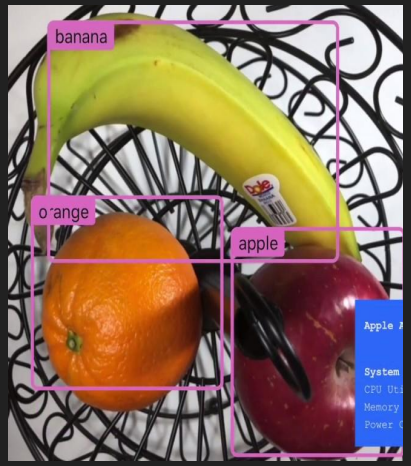
Team 4: ADRAKATTI VIVEK, XIA CHENGUANG, DWIVEDEE LAKSHYAJEET

Index

Problem Formulation

1. Problem Formulation

- Patterns in usage of ingredients
- Recommend ingredients based on other ingredients
- New “smart-stores” like Amazon Go
- Recommend products based on shopping basket



Demo

Exploratory Data Analysis

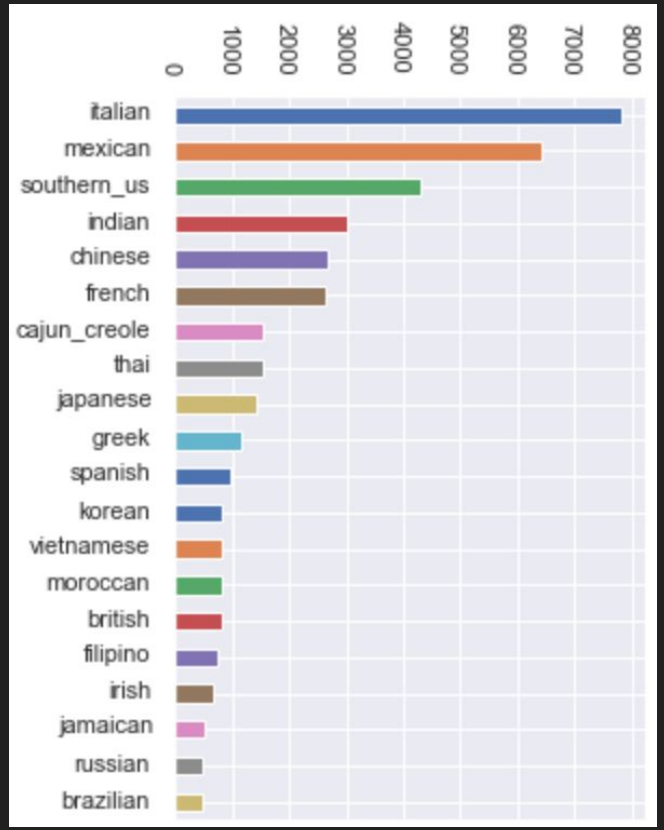
3. EDA - Data Preparation


```
[
  {
    "id": 10259,
    "cuisine": "greek",
    "ingredients": [
      "romaine lettuce",
      "black olives",
      "grape tomatoes",
      "garlic",
      "pepper",
      "purple onion",
      "seasoning",
      "garbanzo beans",
      "feta cheese crumbles"
    ]
  },
  {
    "id": 25693,
    "cuisine": "southern_us",
    "ingredients": [
```


id	cuisine	ingredients	all_ingredients
0	spanish	[mussels, ground black pepper, garlic cloves, ...	mussels;ground black pepper;garlic cloves;saff...
1	mexican	[tomatoes, diced red onions, paprika, salt, co...	tomatoes;diced red onions;paprika;salt;corn to...
2	french	[chicken broth, truffles, pimentos, green pepp...	chicken broth;truffles;pimentos;green pepper;o...
3	chinese	[fresh ginger, sesame oil, frozen peas, cooked...	fresh ginger;sesame oil;frozen peas;cooked ric...
4	italian	[orange peel, cookies, vanilla ice cream, gran...	orange peel;cookies;vanilla ice cream;gran mar...

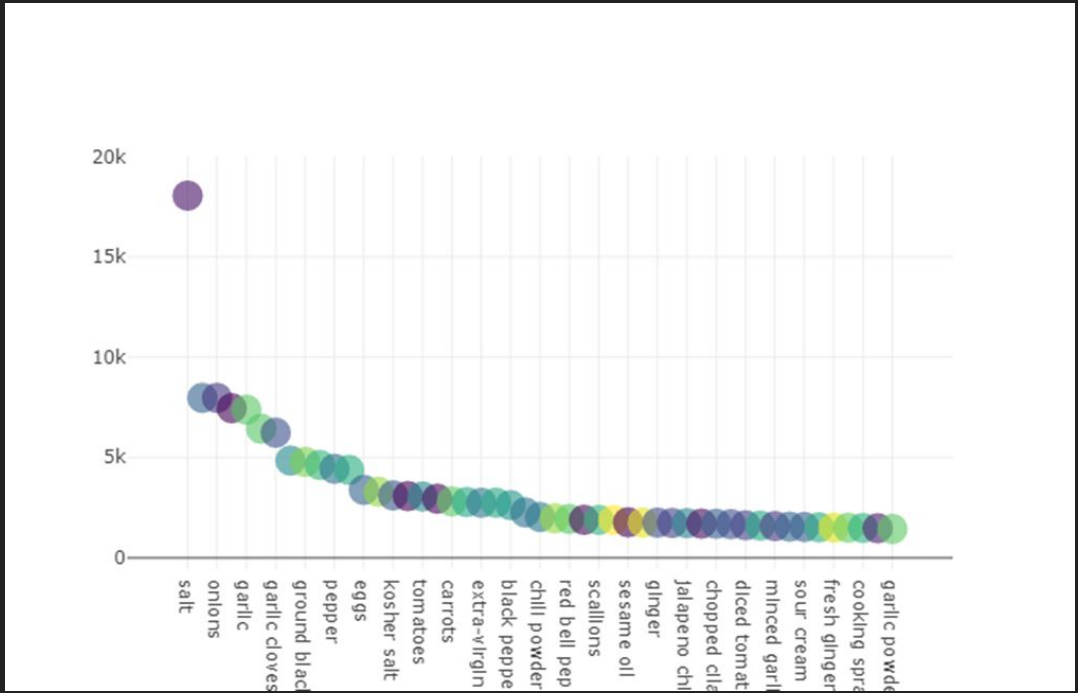
- read_json
- Sort by id
- List of ingredients to “;” separated string
- (39774, 3)
- No null values

3. EDA - Cuisines



- 20 cuisines
- Italian - 7838
- Brazilian - 467

3. EDA - Common Ingredients



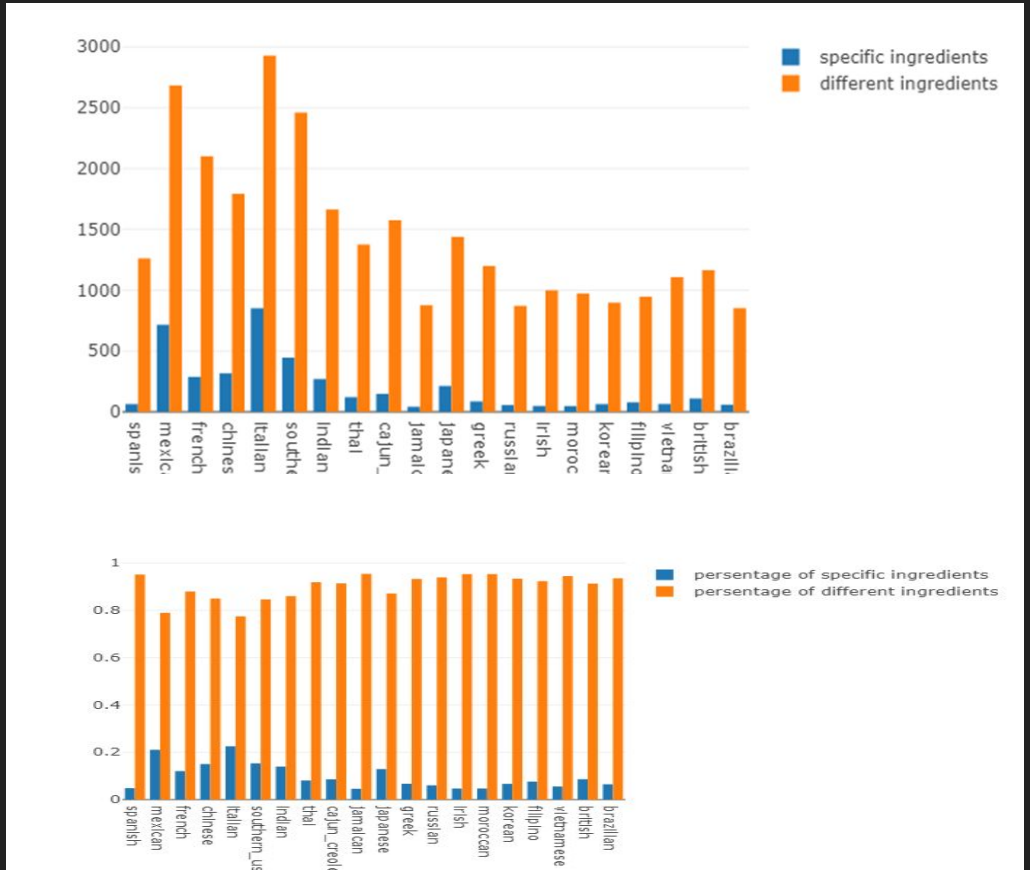
- Most common:
  - Salt
  - Olive Oil
  - Onions
- Repeated Ingredients:
  - Garlic
  - Garlic Cloves

3. EDA - Top 10 Ingredients

	Top 1	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
spanish	salt	olive oil	garlic cloves	extra-virgin olive oil	onions	water	tomatoes	ground black pepper	red bell pepper	pepper
mexican	salt	onions	ground cumin	garlic	olive oil	chili powder	jalapeno chilies	sour cream	avocado	corn tortillas
french	salt	sugar	all-purpose flour	unsalted butter	olive oil	butter	water	large eggs	garlic cloves	ground black pepper
chinese	soy sauce	sesame oil	salt	corn starch	sugar	garlic	water	green onions	vegetable oil	scallions
italian	salt	olive oil	garlic cloves	grated parmesan cheese	garlic	ground black pepper	extra-virgin olive oil	onions	water	butter
southern_us	salt	butter	all-purpose flour	sugar	large eggs	baking powder	water	unsalted butter	milk	buttermilk
indian	salt	onions	garam masala	water	ground turmeric	garlic	cumin seed	ground cumin	vegetable oil	oil
thai	fish sauce	garlic	salt	coconut milk	vegetable oil	soy sauce	sugar	water	garlic cloves	fresh lime juice
cajun_creole	salt	onions	garlic	green bell pepper	butter	olive oil	cayenne pepper	cajun seasoning	all-purpose flour	water
jamaican	salt	onions	water	garlic	ground allspice	pepper	scallions	dried thyme	black pepper	garlic cloves
japanese	soy sauce	salt	mirin	sugar	water	sake	rice vinegar	vegetable oil	scallions	ginger
greek	salt	olive oil	dried oregano	garlic cloves	feta cheese crumbles	extra-virgin olive oil	fresh lemon juice	ground black pepper	garlic	pepper
russian	salt	sugar	onions	all-purpose flour	sour cream	eggs	water	butter	unsalted butter	large eggs
irish	salt	all-purpose flour	butter	onions	potatoes	sugar	baking soda	baking powder	milk	carrots
moroccan	salt	olive oil	ground cumin	onions	garlic cloves	ground cinnamon	water	ground ginger	carrots	paprika
korean	soy sauce	sesame oil	garlic	green onions	sugar	salt	water	sesame seeds	onions	scallions
filipino	salt	garlic	onions	water	soy sauce	pepper	oil	sugar	carrots	ground black pepper
vietnamese	fish sauce	sugar	salt	garlic	water	carrots	soy sauce	shallots	garlic cloves	vegetable oil
british	salt	all-purpose flour	butter	milk	eggs	unsalted butter	sugar	onions	baking powder	large eggs
brazilian	salt	onions	olive oil	lime	water	garlic cloves	garlic	cachaca	sugar	tomatoes

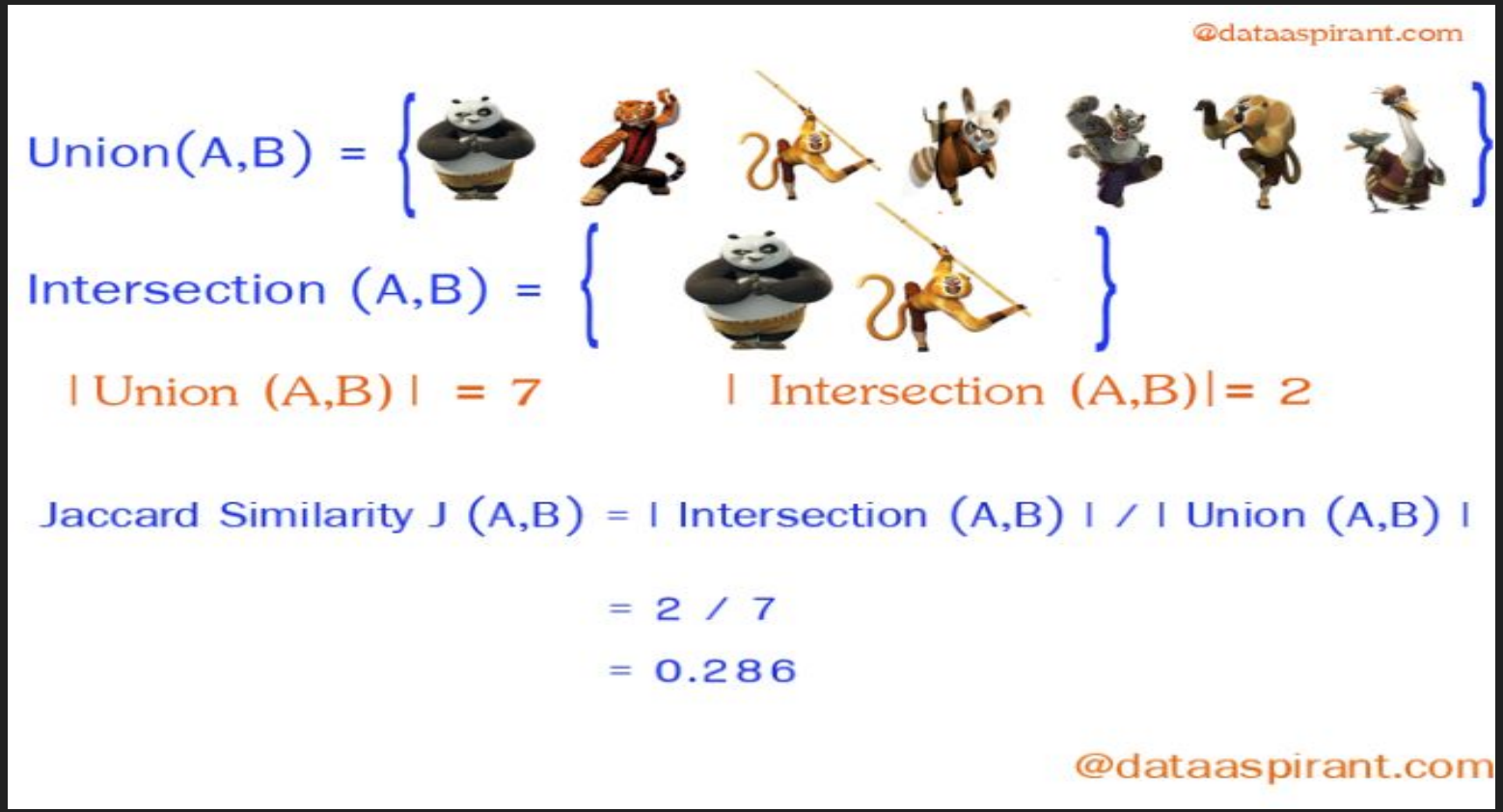
- Salt vs. Soy Sauce
- Olive Oil vs. Sesame Oil
- Mostly spices or oils

3. EDA - Specific / Different Ingredients



3. EDA - Cultural Diffusion by Ingredients

3. EDA - Jaccard similarity

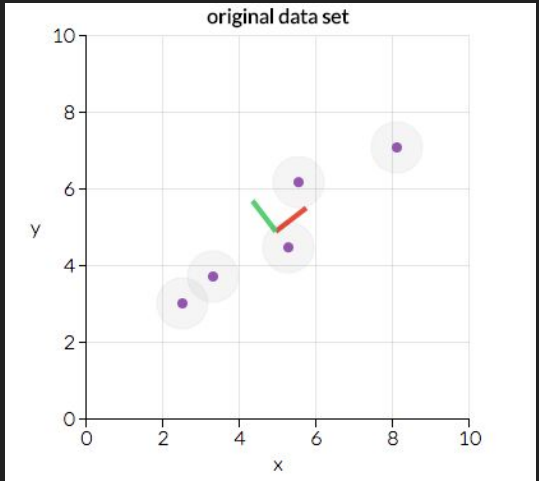


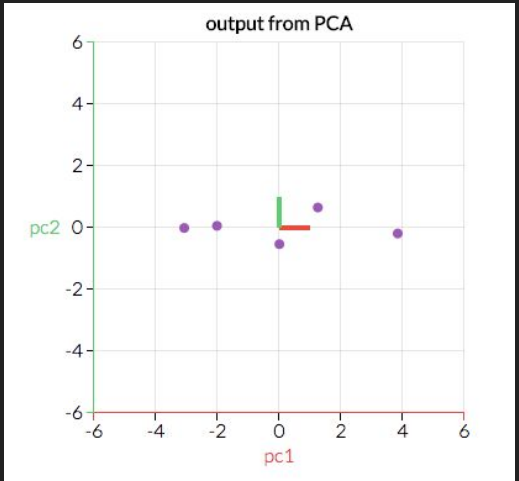
https://dataconomy.com/2015/04/implementing-the-five-most-popular-similarity-measures-in-python/

3. EDA - Principal Component Analysis

Dimensionality Reduction: reducing the dimension of feature space by
1.) Feature elimination 2.) Feature extraction

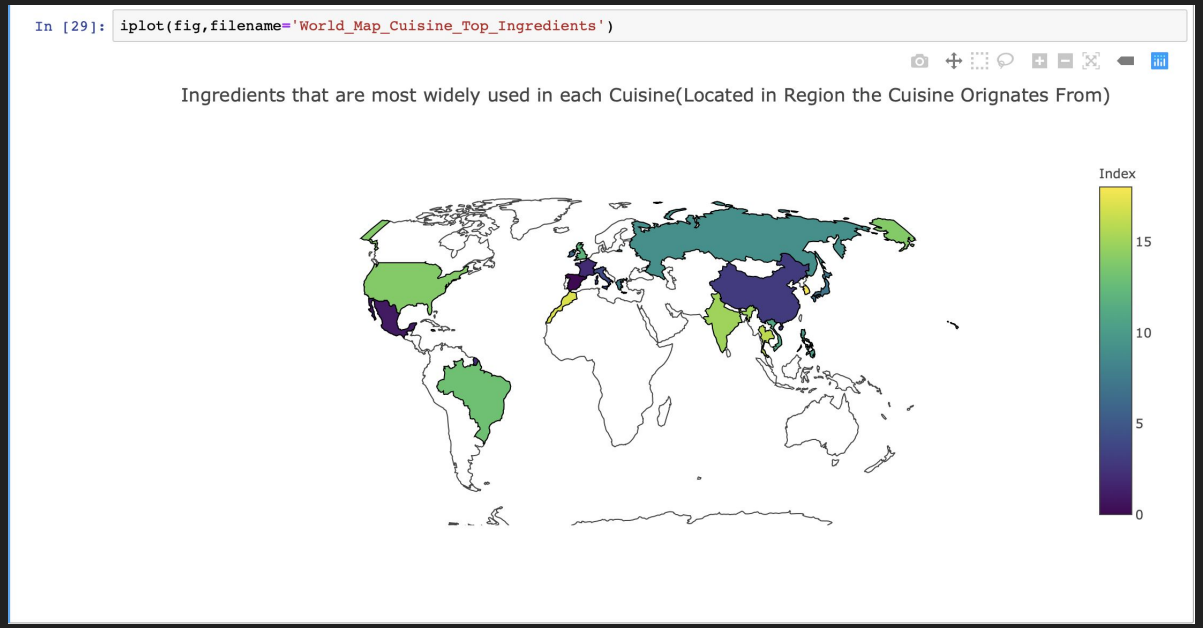
PCA is a technique for feature extraction. It helps to drop the ‘least important’ variables but to retain the most valuable ones.





3. EDA - Choropleth

- Hover over country to show top 3 ingredients that “define” the local cuisine
- By finding most common ingredients in every cuisine
  - Manual filtering to remove universal ingredients



3. EDA - Choropleth(Demo)

Cuisine Prediction

4. Cuisine Prediction - Encoding

Count Vectoriser

	Jumps	The	brown	dog	fox	lazy	over	quick	the
Doc1	0	1	1	0	1	0	0	1	0
Doc2	1	0	0	1	0	1	1	0	1

- ML models take numbers
- One-hot encoding for each ingredient

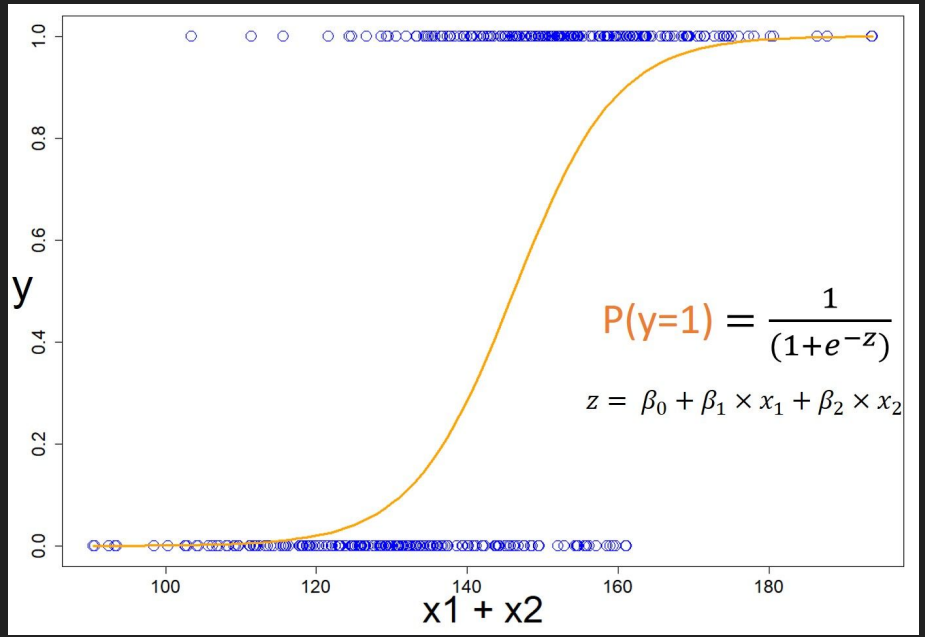
Label Encoder

CAT73	label_encoded
A	1
A	1
C	3
B	2
A	1
C	3
B	2

Assign Number to each cuisine

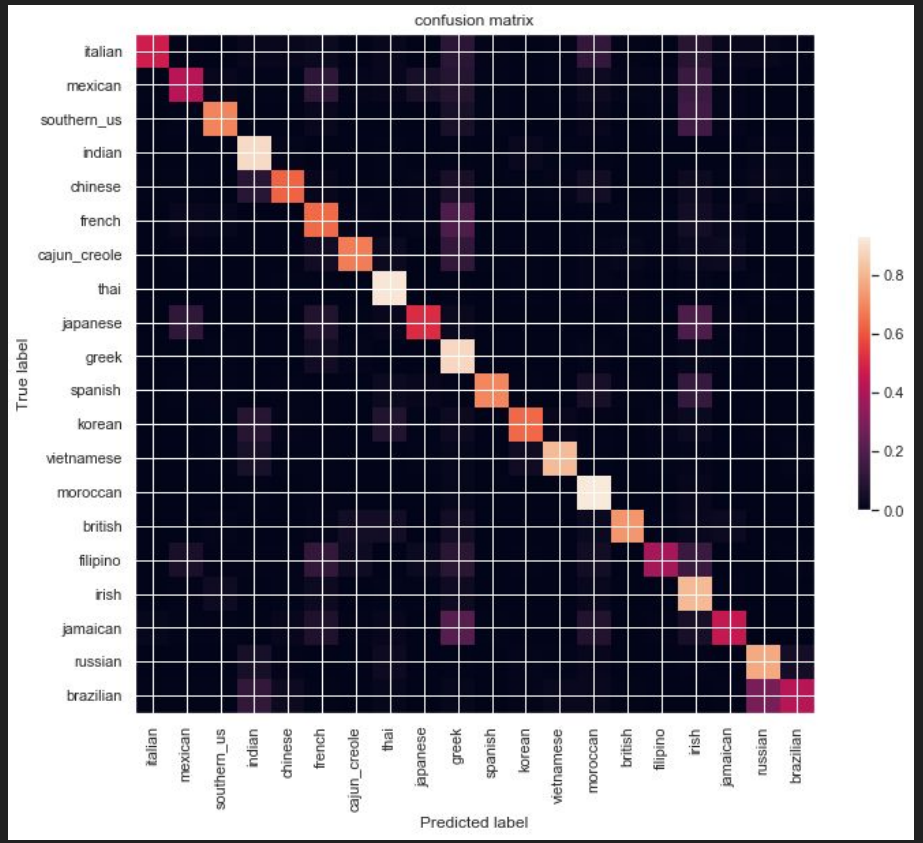
5. Cuisine Prediction - Logistic Regression

- Fit an S-shaped curve between 2 classes
- Output - Probability of Y based on X



4. Cuisine Prediction - Confusion Matrix

- Moroccan, Thai, Indian are well predicted
- Greek misidentified most
- Accuracy - 78.4%



Ingredient Recommendation

5. Ingredient Recommendation - Naive Method

- Count co-occurrence
- Sort by highest co-occurrence
- Problem:
  - Most common ingredients in dataset

	a	b	a_count	b_count	cooc
40197	condensed milk	sugar	74	6434	33
84291	condensed milk	water	74	7457	25
40200	condensed milk	eggs	74	3388	19
71303	condensed milk	vanilla extract	74	1298	16
71301	condensed milk	egg yolks	74	542	13
40202	condensed milk	evaporated milk	74	208	13
47904	condensed milk	milk	74	2263	13
166595	condensed milk	whole milk	74	764	9
86943	condensed milk	salt	74	18049	9
79010	condensed milk	unsalted butter	74	2782	7

5. Ingredient Recommendation - Using PMI

Pointwise Mutual Information

$$PMI(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)}$$

- Compare P(X and Y together) against P(X and Y not together)
- Ingredients can be common, not occur together

5. Ingredient Recommendation - PMI Analysis

Low Count | Very High PMI

a	b	a_count	b_count	cooc	pmi
light cream or half and half	lipton green tea bag	1	1	1	7.053491
frozen basil	red wine vinaigrette	1	1	1	7.053491
mazolaâ® chicken flavor bouillon powder	spice islands bay leaves	1	1	1	7.053491
miswa	pork heart	1	1	1	7.053491
chipped beef	chourico	1	1	1	7.053491
mccormick taco seasoning	tomato sauce low sodium	1	1	1	7.053491
gluten free barbecue sauce	rub seasoning	1	1	1	7.053491
chuck short ribs	prune juice	1	1	1	7.053491
hawaiian salt	raw buckwheat groats	1	1	1	7.053491
raw buckwheat groats	sliced mango	1	1	1	7.053491

- Barely occur, giving high PMI

Average Count | High PMI

a	b	a_count	b_count	cooc	pmi
herdez salsa casera	herdez salsa verde	5	6	3	4.750905
black treacle	porridge oats	7	5	3	4.596755
kewra water	stone flower	6	6	3	4.568584
sazon seasoning	sofrito	5	5	2	4.527762
mo hanh	vegan mayonnaise	5	6	2	4.345440
chinese rose wine	maltose	5	9	3	4.345440
rye whiskey	twists	6	5	2	4.345440
johnsonville andouille	red goldâ® diced tomatoes	6	5	2	4.345440
black rice vinegar	chinese sesame paste	9	5	3	4.345440
bertolli vineyard premium collect marinara wit...	clams, well scrub	5	6	2	4.345440

- Branded ingredients skew data

5. Ingredient Recommendation - PMI Analysis

Any Count | Negative PMI

a	b	a_count	b_count	cooc	pmi
pepper	vanilla extract	4438	1298	2	-7.819901
sesame oil	unsalted butter	1773	2782	2	-7.664716
garlic	powdered sugar	7380	501	2	-7.376497
cooking spray	soy sauce	1490	3296	3	-7.254893
garlic	large garlic cloves	7380	873	4	-7.238680
diced tomatoes	sesame oil	1624	1773	2	-7.126438
buttermilk	soy sauce	863	3296	2	-7.114242
grated parmesan cheese	soy sauce	1886	3296	5	-6.979750
confectioners sugar	garlic cloves	395	6237	2	-6.970503
extra-virgin olive oil	vanilla extract	2747	1298	3	-6.934742

- Occur separately more often than together

Count > 30 | PMI ~ 2

a	b	a_count	b_count	cooc	pmi
brown cardamom	green cardamom	40	86	27	2.206101
gari	wasabi	50	32	12	2.160638
dried bonito flakes	konbu	42	66	18	2.016538
bonito flakes	konbu	38	66	16	1.998838
asafoetida powder	fresh curry leaves	33	69	13	1.887826
sushi rice	wasabi	81	32	14	1.832363
coffee granules	coffee liqueur	34	30	5	1.735371
condensed cream of chicken soup	condensed cream of mushroom soup	62	55	16	1.691612
brown cardamom	mace	40	81	15	1.678212
asafoetida powder	dried red chile peppers	33	103	15	1.630304

- True relationships

5. Ingredient Recommendation - Final Steps

- Matrix Factorisation
- Singular Value Decomposition
- Cosine Similarity

```
result_ings, ing_scores = display_most_similar('beans', 20)
```

```
- Most similar to 'beans'
  . chihuahua cheese : 1.00
  . feta cheese crumbles : 0.54
  . red wine vinaigrette : 0.54
  . frozen basil : 0.53
  . whole wheat penne : 0.52
  . merguez sausage : 0.49
  . roasted tomatoes : 0.49
  . cornbread mix : 0.49
  . full fat sour cream : 0.49
  . chipotle : 0.49
  . gluten free cooking spray : 0.49
  . baby broccoli : 0.49
  . hurst family harvest chipotle lime black bean soup mix : 0.48
  . chili seasoning : 0.48
  . kraft sharp cheddar cheese : 0.48
  . crystal farms reduced fat shredded marble jack cheese : 0.48
  . gluten free corn tortillas : 0.48
  . cottage cheese : 0.47
  . cooked beetroot : 0.47
  . ground turkey : 0.46
```

Image Detection

6. Using GCV for Image Detection

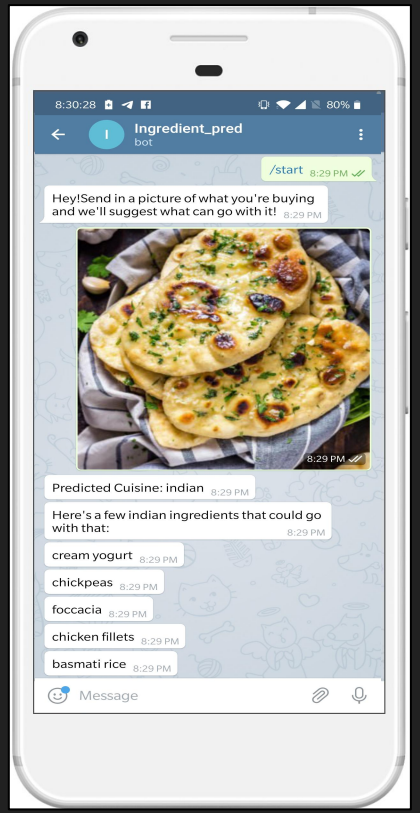
- Used Google Cloud Vision API
- Using GCV to generate tags for image
- Check if tags match up with existing ingredients
- Use ingredient recommendation model to predict recommended ingredients

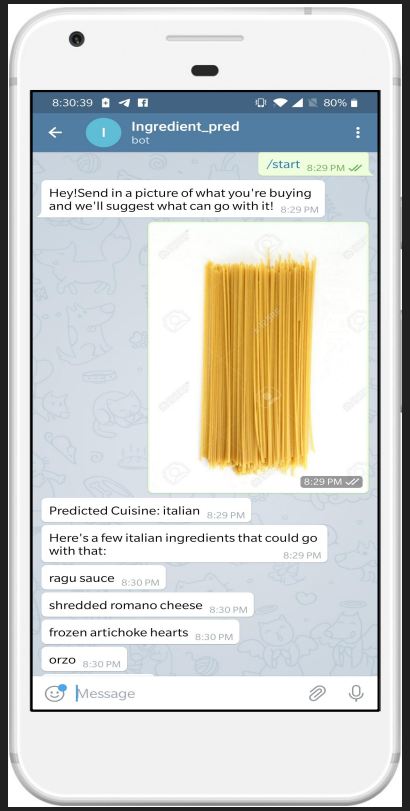
Deployment

7. Using Telegram for Deployment

- Use Telegram Bot API to make a telegram bot
- Provides ease of use and wide reach
- Telegram has 200 million+ users(Mar 2018)
- This wide reach will allow for easy deployment
- Simple, direct and familiar UI for users

7. Using Telegram for Deployment





Further Improvements

8. Further Improvements

- Model
  - Using n-grams for branded ingredient and duplicated ingredient problem (Ginger vs. Fresh Ginger)
  - Predict ingredient based on desired cuisine
- Prediction for more than a single ingredient
- Independent application instead of telegram bot
- Given calorie data, calculate total calories of all ingredients

Notes

- Telegram Bot is working for three ingredients currently as POC
  - Spaghetti, Beans, and Naan
- Can be extended to all ingredients in the data in the future
- Cuisine prediction and ingredient recommendation systems made in the Jupyter Notebooks have been exported to python files to be used by flask server
- Models folder contains Pickle files of models and data needed to run system on flask.