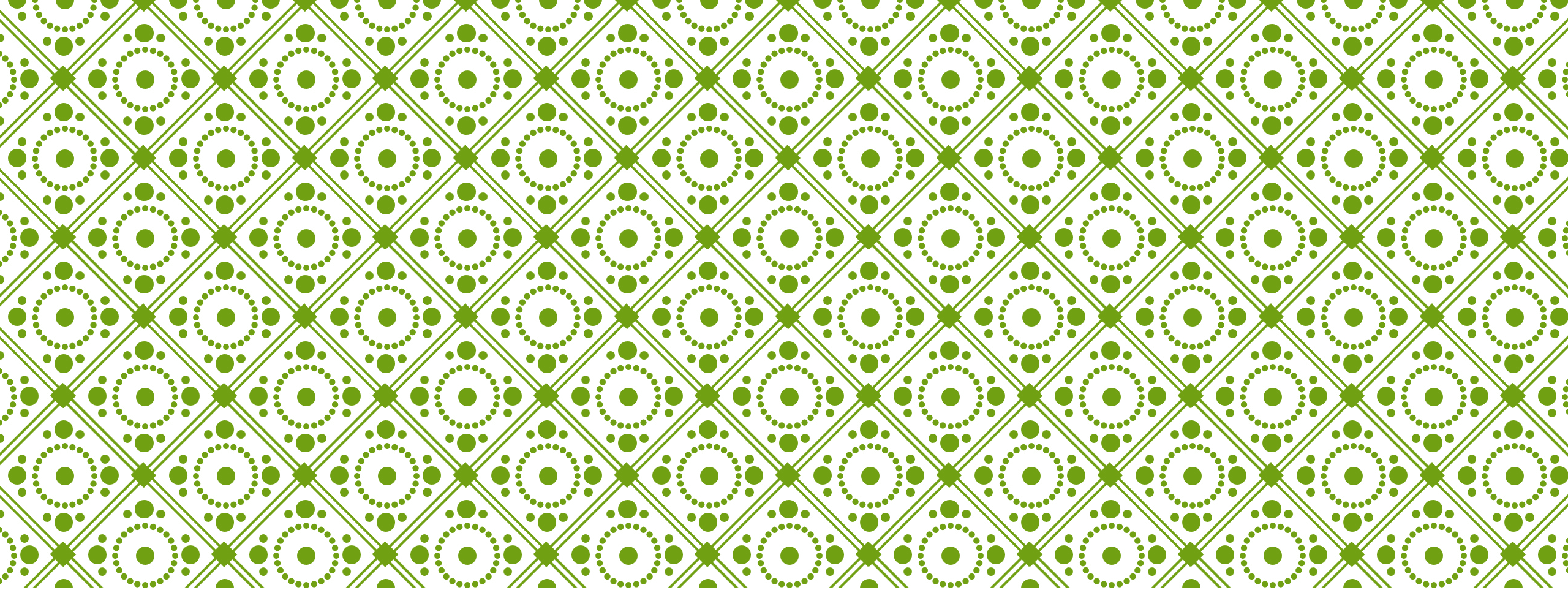
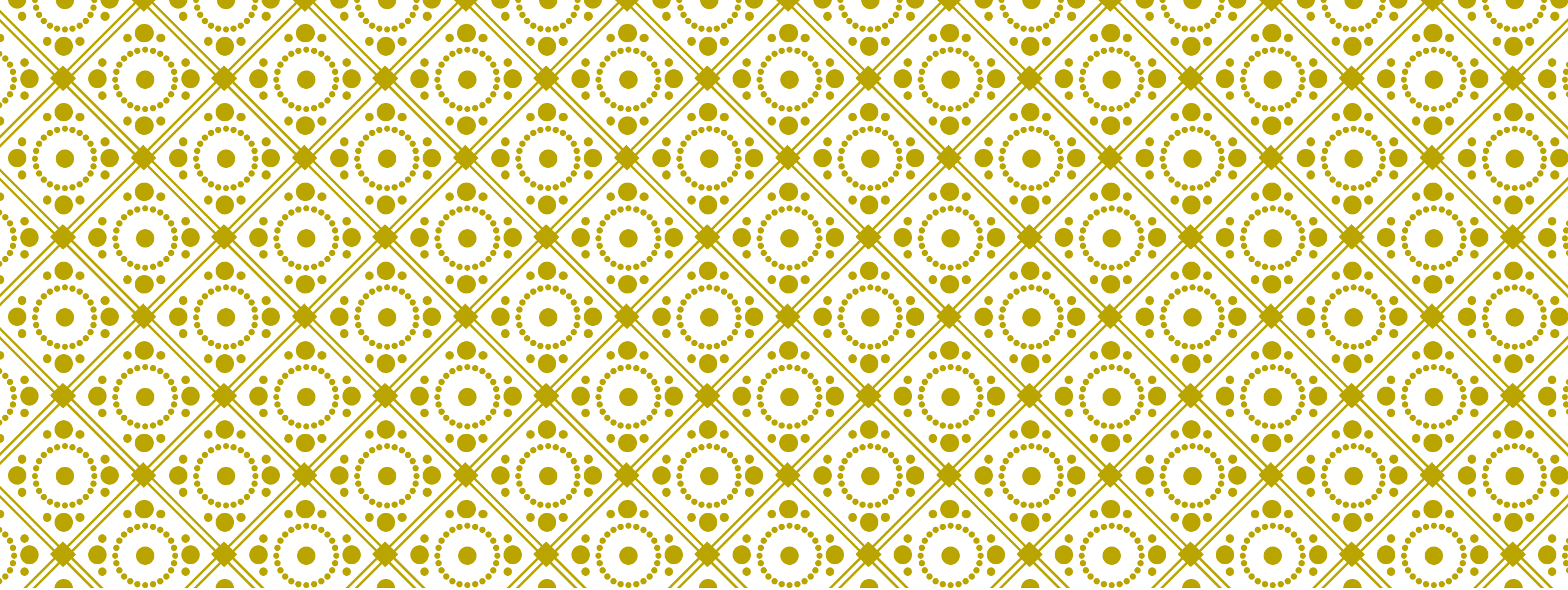


FINAL PROJECT PRESENTATION

Benny Y. Pratama
JC Data Science Batch 1

OUTLINES

- Objectives
- Data Analysis
- Data Preprocessing
- Prediction and Evaluation

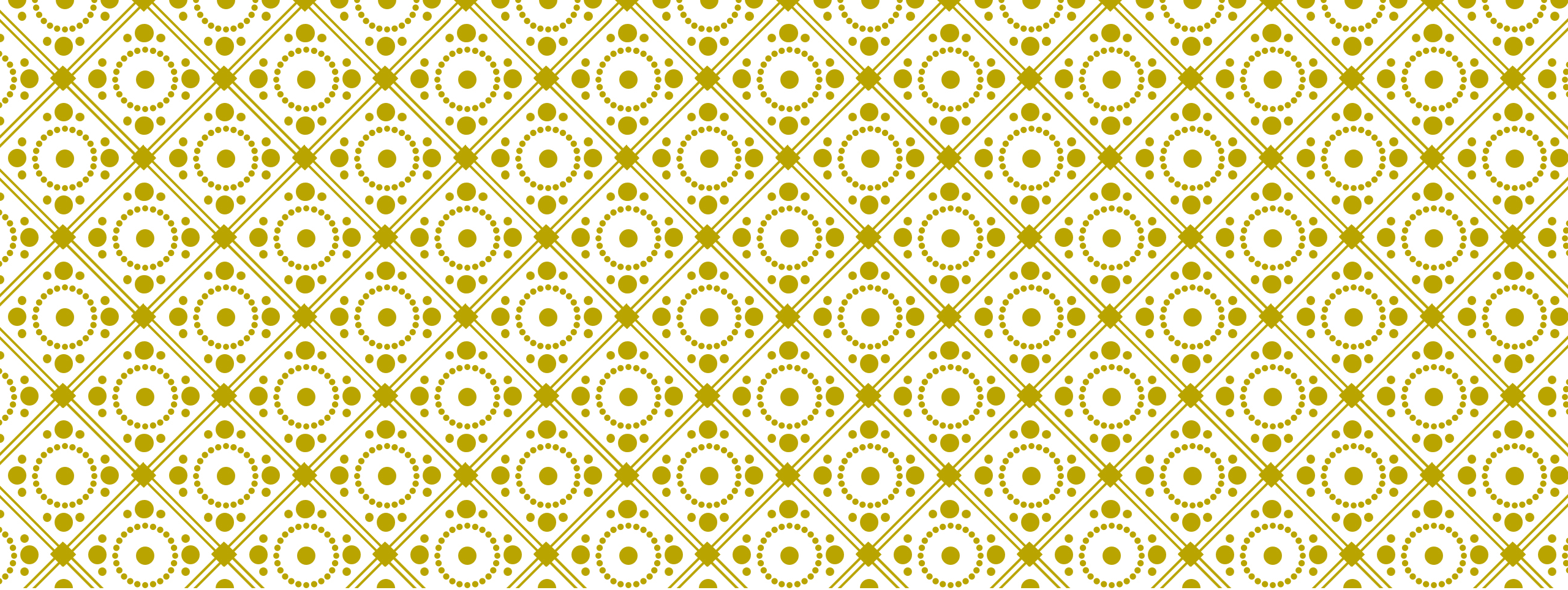


OBJECTIVES

OBJECTIVES – WHAT'S COOKING?

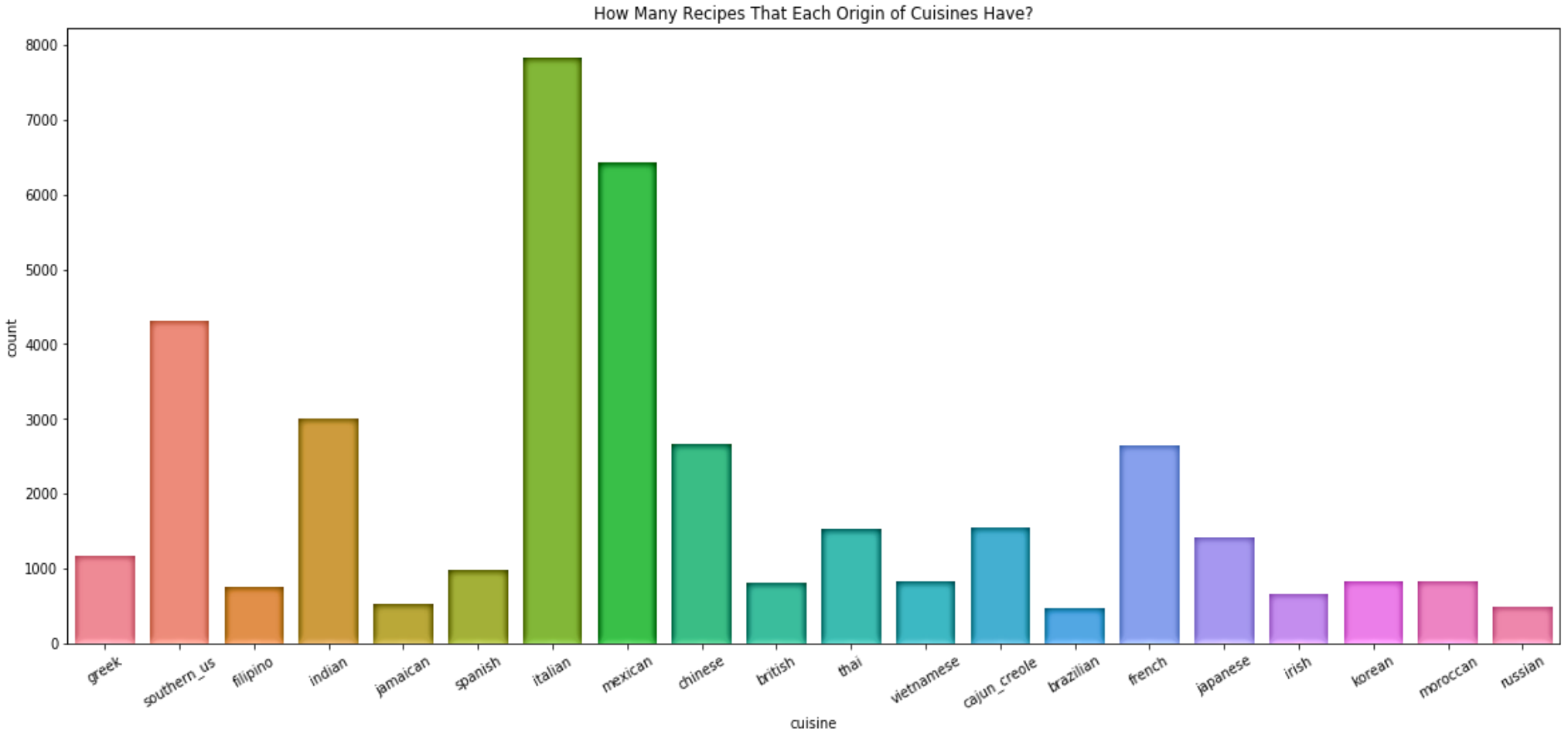
To predict the origin of a recipe based on the ingredients provided.

Source : Kaggle (https://www.kaggle.com/c/whats-cooking-kernels-only, provided by Yummly - https://www.yummly.com/)

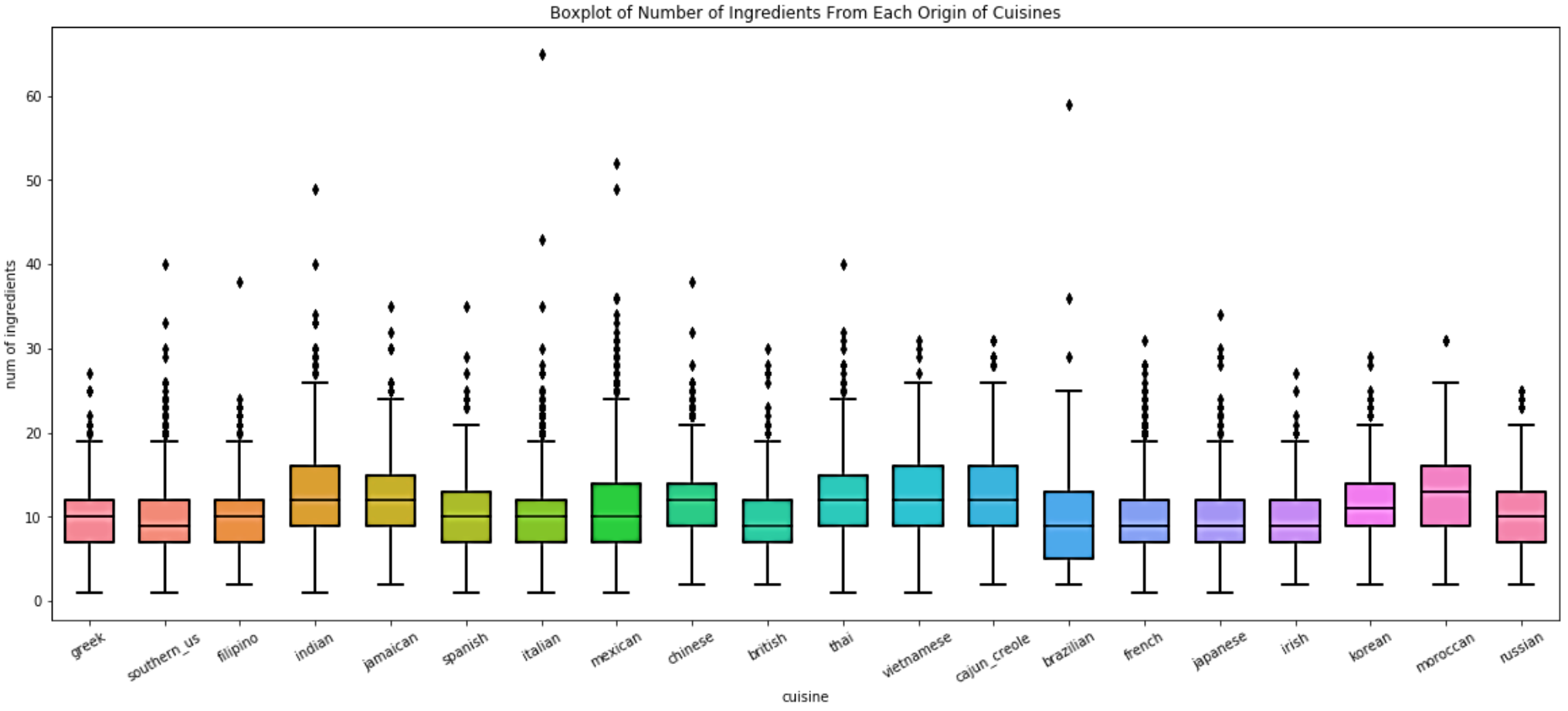


DATA ANALYSIS

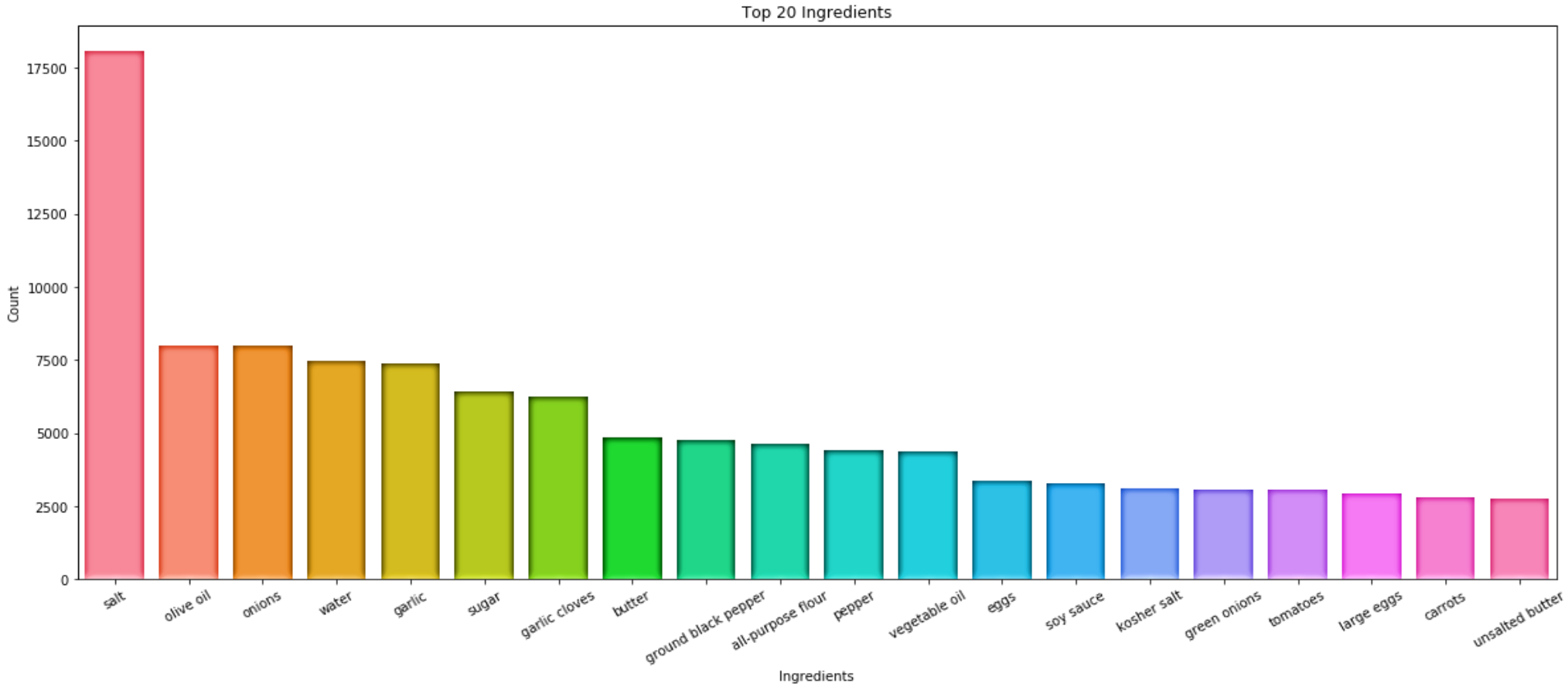
DATA ANALYSIS – OVERALL RECIPE



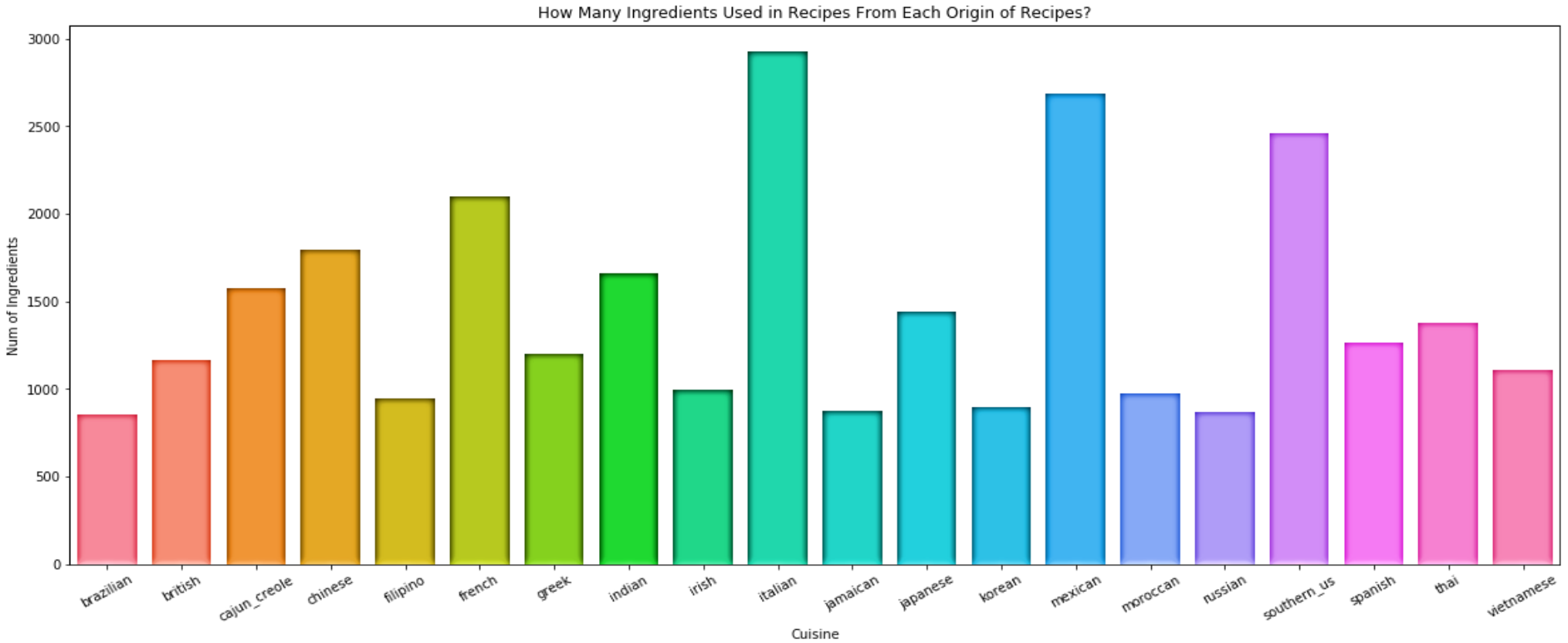
DATA ANALYSIS – BOXPLOT



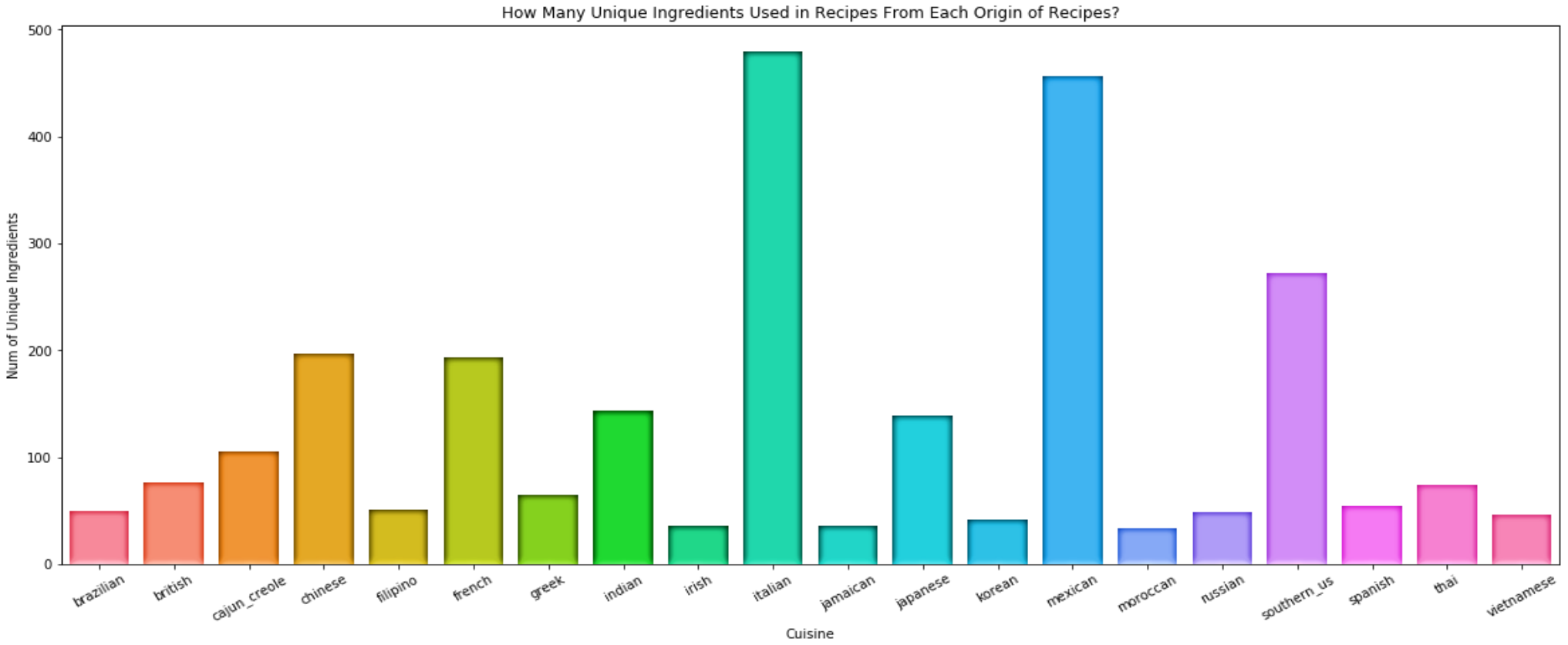
DATA ANALYSIS – TOP 20 INGREDIENTS



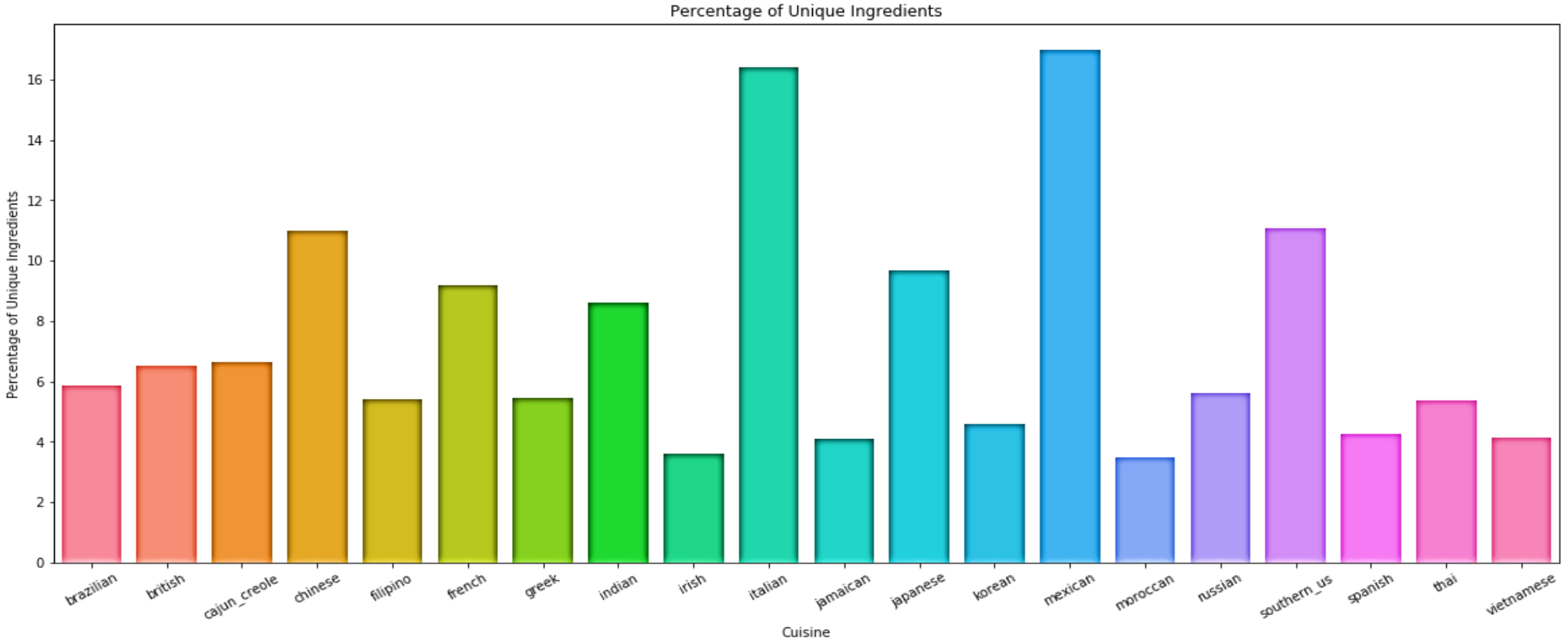
DATA ANALYSIS - HOW MANY INGREDIENTS USED IN RECIPES FROM EACH ORIGIN OF RECIPES?



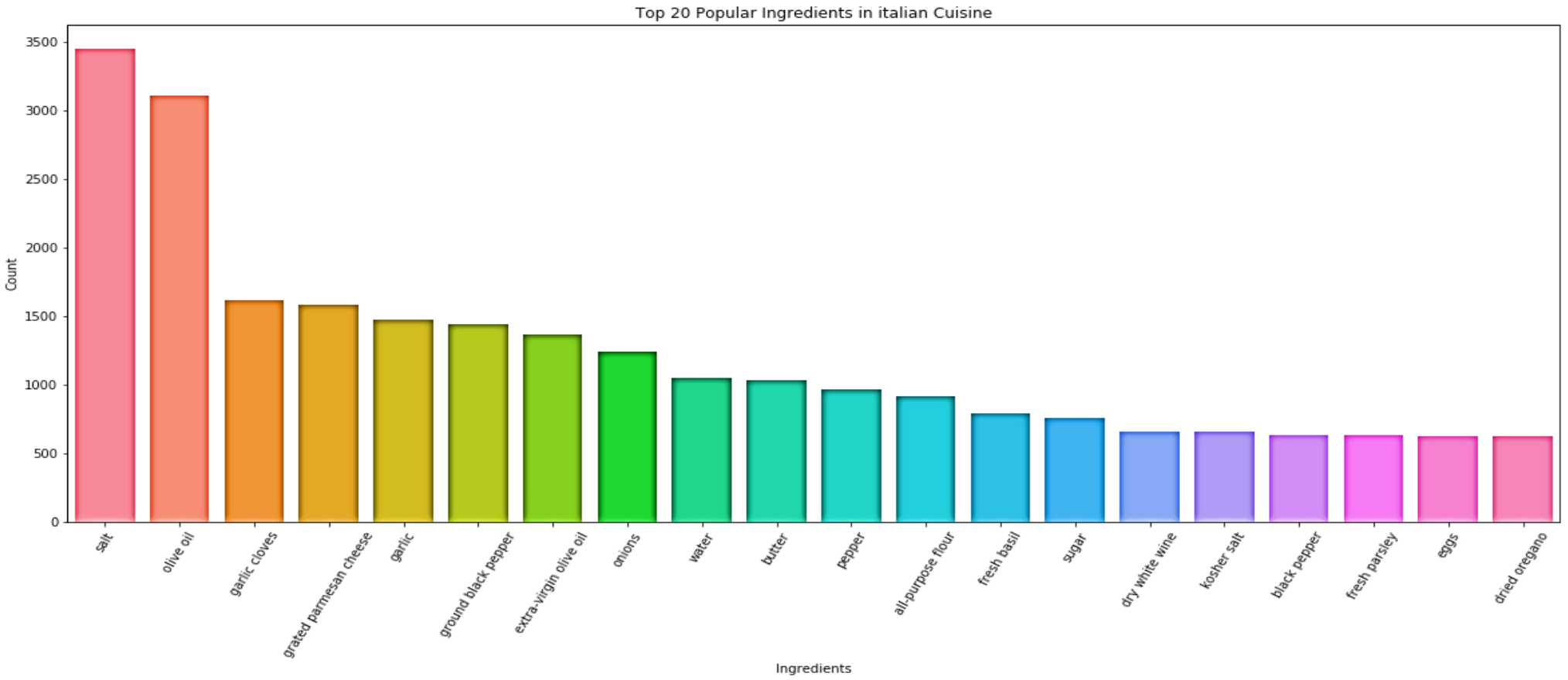
DATA ANALYSIS - HOW MANY UNIQUE INGREDIENTS USED IN RECIPES FROM EACH ORIGIN OF RECIPES?



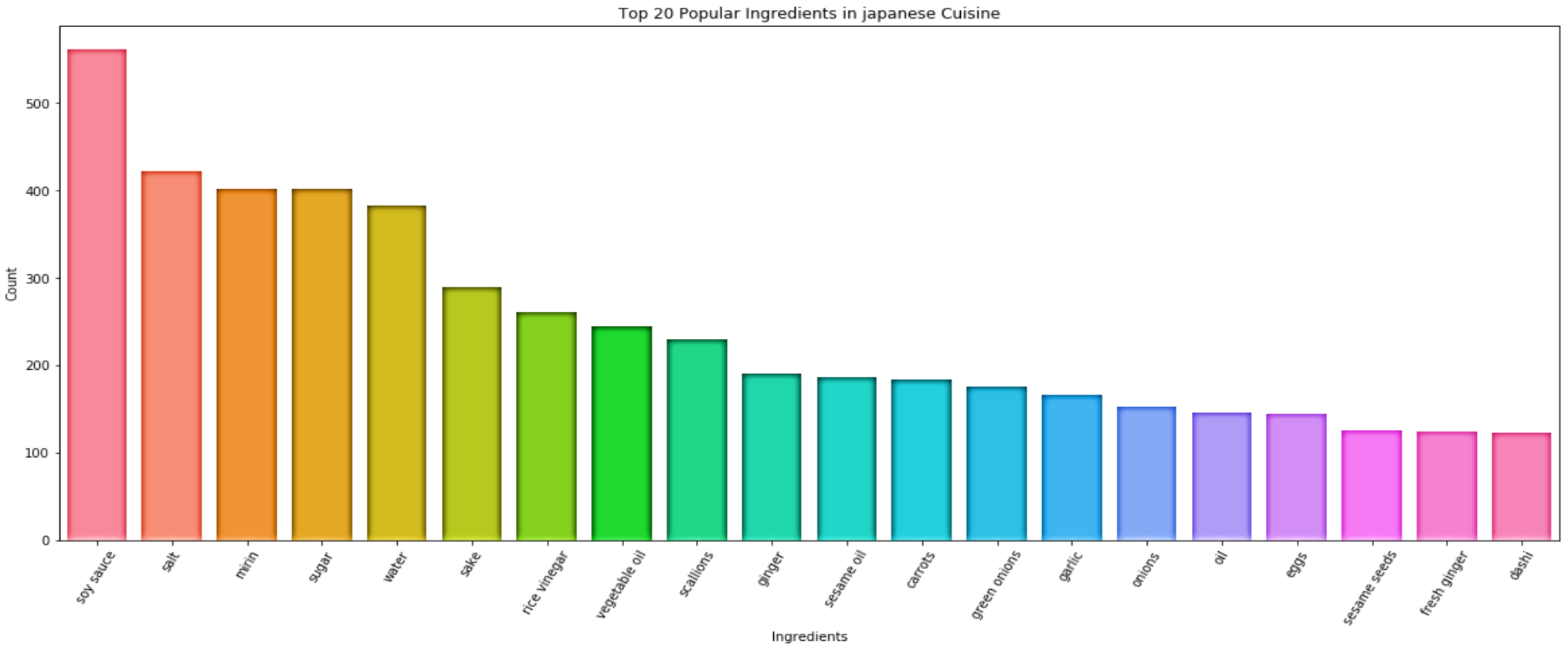
DATA ANALYSIS - PERCENTAGE OF UNIQUE INGREDIENTS



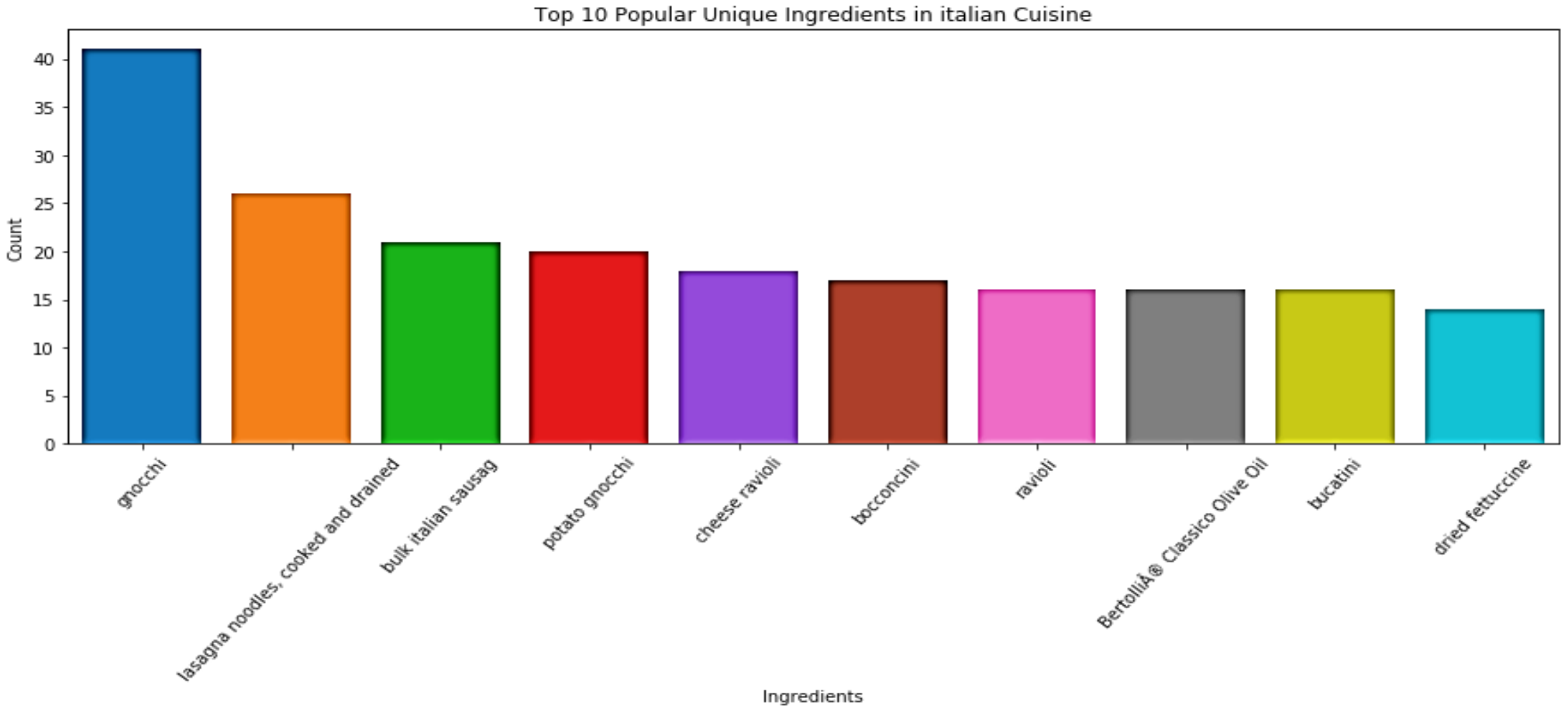
DATA ANALYSIS – TOP 20 OF INGREDIENTS FROM SOME ORIGIN OF RECIPES (ITALY)



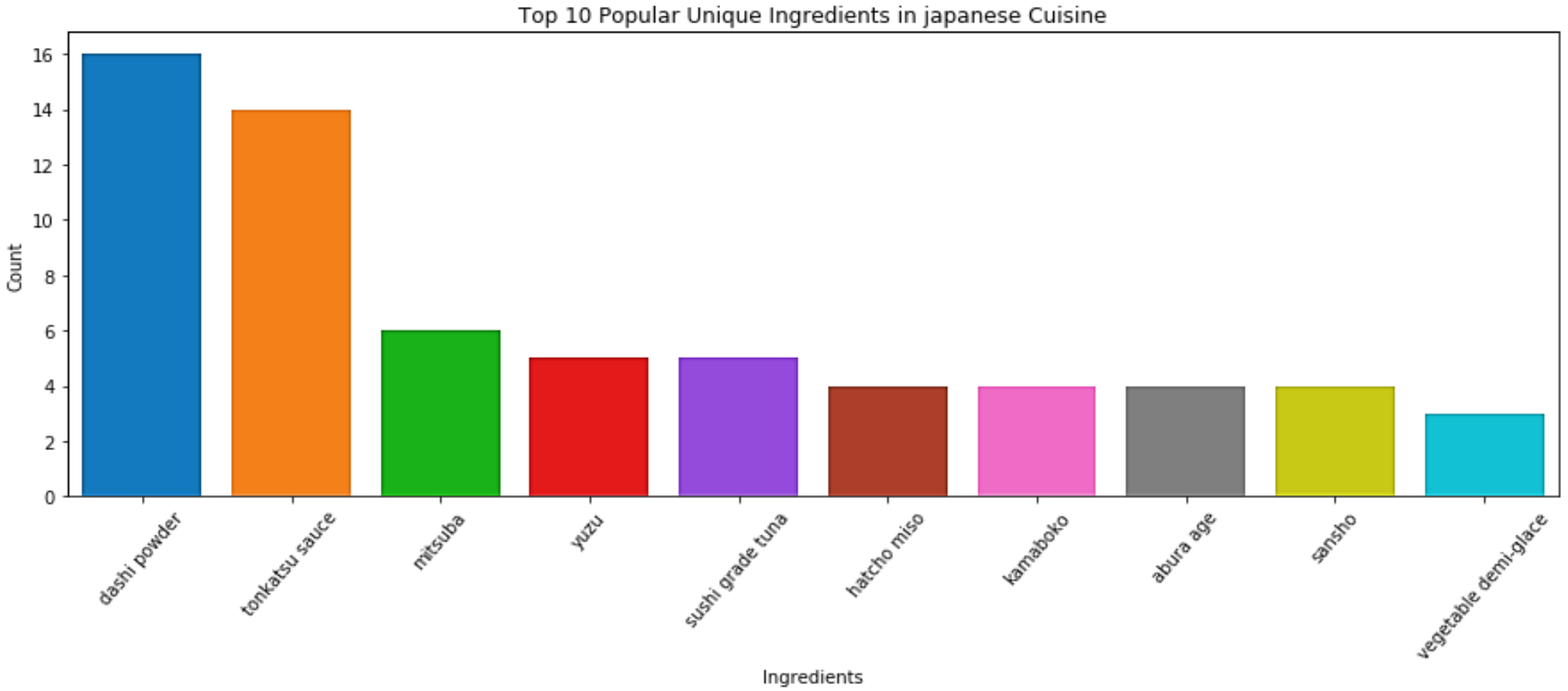
DATA ANALYSIS – TOP 20 OF INGREDIENTS FROM SOME ORIGIN OF RECIPES (JAPAN)

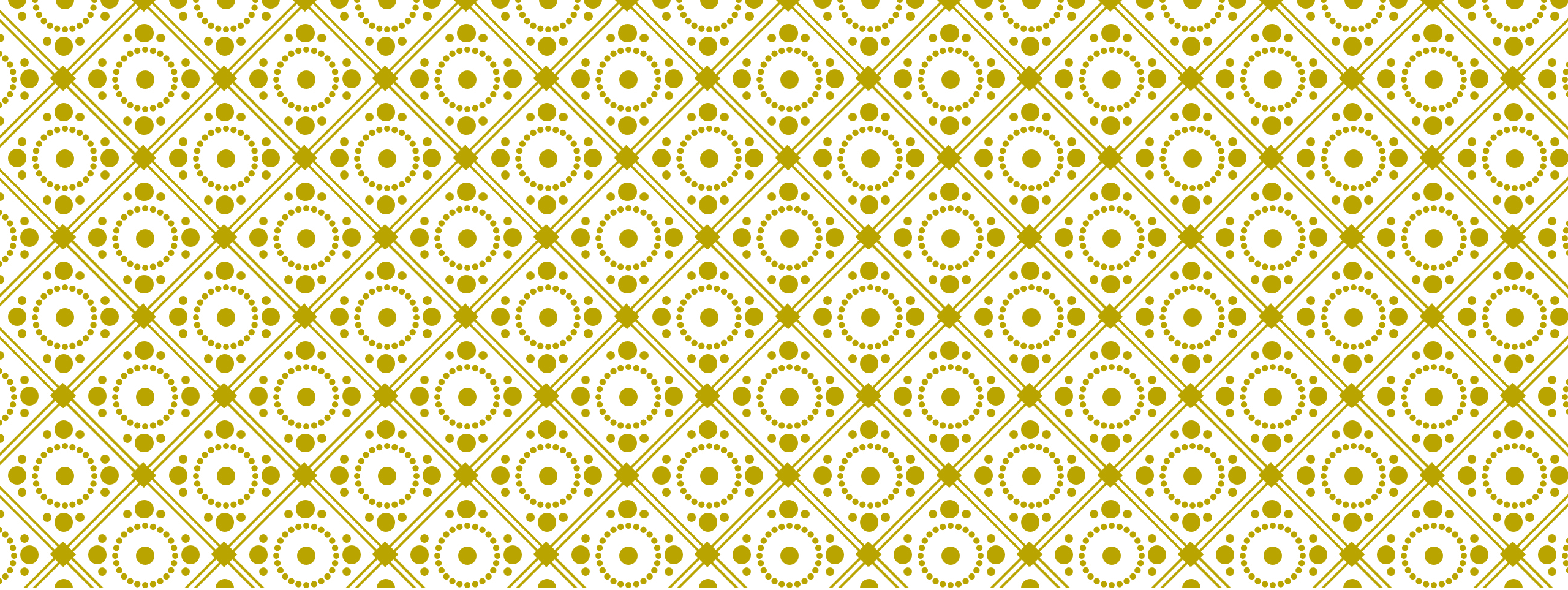


DATA ANALYSIS – TOP 20 OF UNIQUE INGREDIENTS FROM SOME ORIGIN OF RECIPES (ITALY)



DATA ANALYSIS – TOP 20 OF UNIQUE INGREDIENTS FROM SOME ORIGIN OF RECIPES (JAPAN)





DATA PREPROCESSING

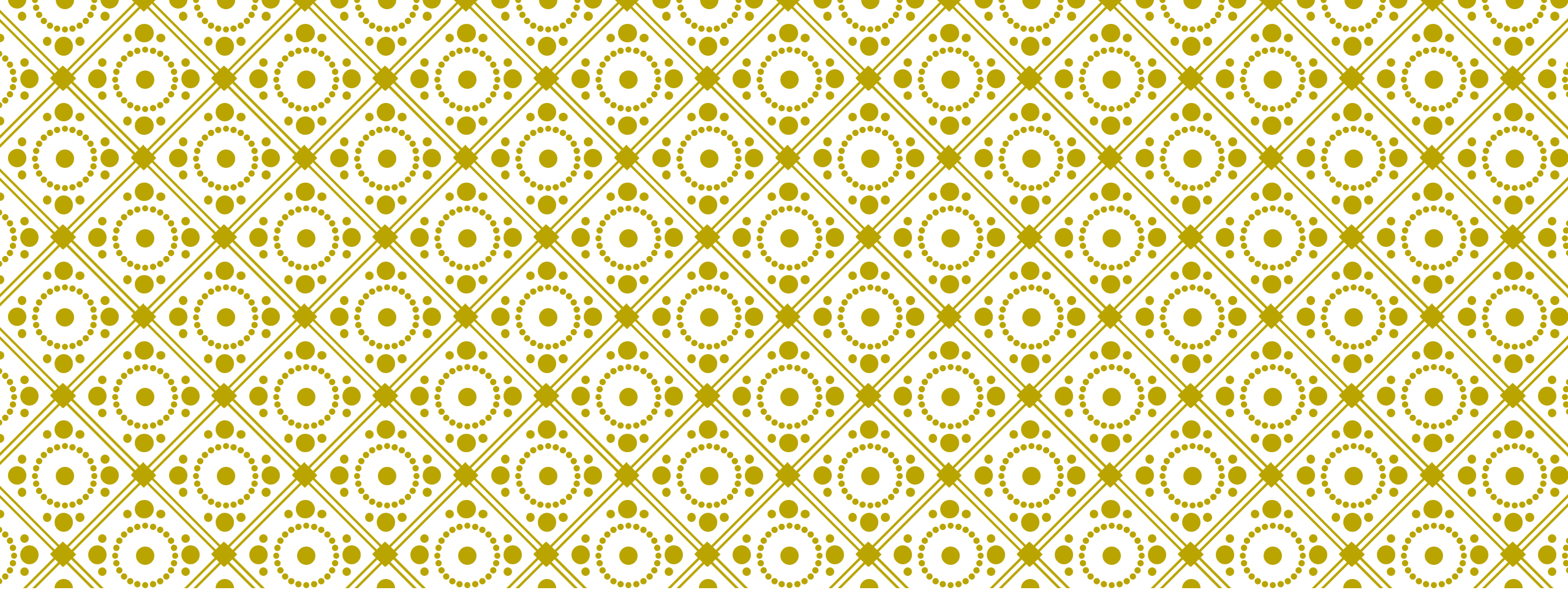
DATA PREPROCESSING - STEPS

- Tokenization data
- Split data into training data and testing data
- Implement Doc2Vec model for word embedding
- Train model using training data
- Set training data and testing data as learning vector

DOC2VEC IN A GLANCE

- An improvement to Word2Vec
- Generally used to find relation among words
- Implement CBOW, Skip-Gram, PV-DM, and PV-DBOW algorithms

Source : https://medium.com/scaleabout/a-gentle-introduction-to-doc2vec-db3e8c0cce5e



PREDICTION AND EVALUATION

MODEL

- Implement LinearSVC model
- It is commonly used for predicting text data

MOMENT OF TRUTH

Classification Report :

	precision	recall	f1-score	support
brazilian	0.44	0.09	0.14	141
british	0.37	0.10	0.16	201
cajun_creole	0.57	0.43	0.49	394
chinese	0.54	0.70	0.61	664
filipino	0.32	0.13	0.18	179
french	0.45	0.42	0.43	618
greek	0.67	0.38	0.49	300
indian	0.69	0.76	0.73	737
irish	0.42	0.08	0.14	182
italian	0.60	0.81	0.69	1990
jamaican	0.62	0.16	0.25	126
japanese	0.66	0.45	0.54	369
korean	0.48	0.40	0.44	210
mexican	0.65	0.84	0.74	1574
moroccan	0.54	0.36	0.43	188
russian	0.47	0.07	0.12	131
southern_us	0.53	0.62	0.57	1075
spanish	0.58	0.13	0.21	239
thai	0.53	0.43	0.47	402
vietnamese	0.37	0.08	0.14	224
micro avg	0.58	0.58	0.58	9944
macro avg	0.53	0.37	0.40	9944
weighted avg	0.57	0.58	0.55	9944

Average Precision : 0.5252790515229597
Average Recall : 0.3720309446978258
Accuracy : 0.584875301689461
Average F1 Score 0.3978736948619979
Matthews Corrcoef : 0.5309195337419191