So far:  
Used apify to scrape AllRecipes website

Only 3890? Data points, will need to scrape more websites later

Preprocessing, need to split ingredients on comma

Eggs/egg yolks/egg whites are an ingredient, deal with separately

New column for other information “room temperature”

LOAD CSV WITH HEADERS FROM 'file:///Ingredient\_Split.csv' AS row

UNWIND RANGE(1, 31) AS x

WITH DISTINCT row, x

MERGE (i:Ingredient { name: row['new\_Ingredient ' + x + '\_Name'] })

WITH DISTINCT row, x

MATCH (i:Ingredient)

MERGE (r:Recipe {recipeId: i.recipeId }) ON CREATE SET r.recipeId = i.recipeId

CREATE (r)-[:CONTAINS]->(i)

Current issues with webscraping

* Trying to find api rather than scraping across each webpage, however, in recipe websites ingredient data isn’t sent as a request every time it is loaded, it is just a html element

Websites:

https://medium.com/analytics-vidhya/python-selenium-web-scraping-in-eight-steps-7d33b263f399#id\_token=eyJhbGciOiJSUzI1NiIsImtpZCI6Ijg1YmE5MzEzZmQ3YTdkNGFmYTg0ODg0YWJjYzg0MDMwMDQzNjMxODAiLCJ0eXAiOiJKV1QifQ..eHCVKimt\_kHQgquGNkg7Hn3qvuzKr0u-6FUGwkCRrpZUO8FieS7WcmZWesDHhrHZY9emzt0HO4k5vPQNt6QX7Kgr-\_Yp6w8J22Q82r1JILj3j\_bnmFUAY7w9fQE1tsuwRrpPGUjLcgThgcRvxVSwQT\_bYispU1vVCuO13G8zXuldB5pkxrAGb7MwBZYYrWOd9gRUEyS\_AWqSb8wkcUsZMZdMeFs44v-zwvSsOuxHq4WOPN9xwTmNqxbKb0VZcPcxPW2X6TlbQxPBtWKpqdY1OsvpU9vlECYikdDdSQAvejy-rJh9zbrr4qO7x2tUoBafg2Or84zk2ytwqBL\_tfDQmg