GRAPH ANALYSIS - MOVIES AND USERS

EXPLORING RELATIONSHIPS IN MOVIE DATA

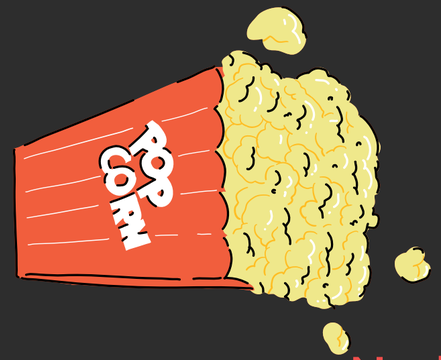
ŁUKASZ TUREK, KRZYSZTOF OSTRZYCKI

BUSINESS DOMAIN

Descritpion: This project analyzes relationships between users, movies, actors, directors, and genres using a graph-based approach. The interconnected data structure reveals patterns and insights into user preferences and movie dynamics.

Goal: The goal is to uncover trends in ratings, genre preferences, and industry relationships.

NODES AND RELATIONSHIPS



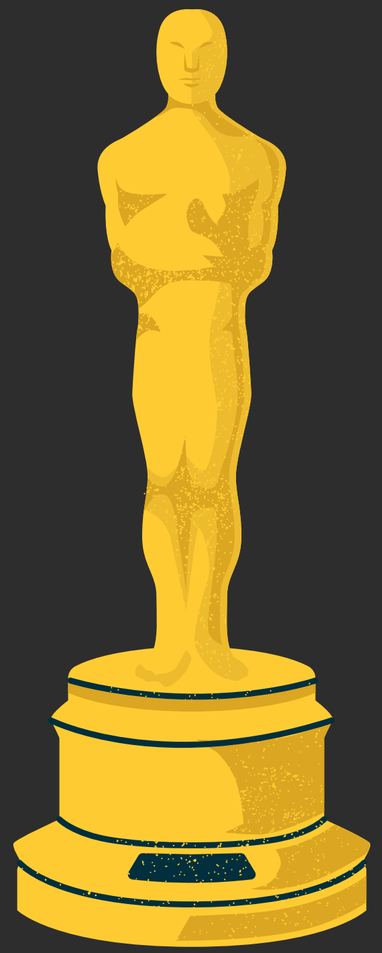
Relationships

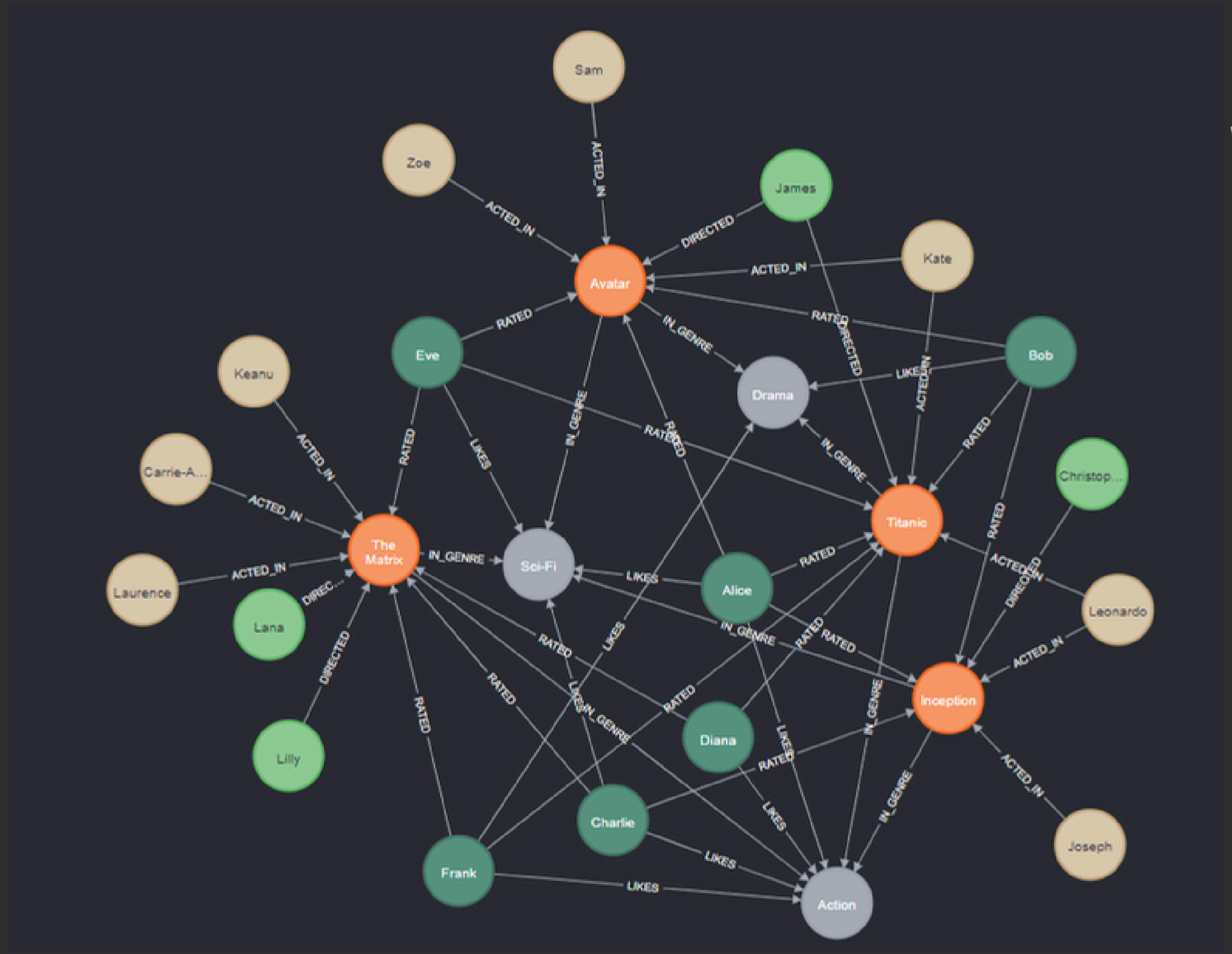
RATED: Connects a user to a movie with a rating attribute.
Example: (User1)-[:RATED {rating: 4}]->(Movie1)
LIKES: Connects a user to genres they prefer.
Example: (User1)-[:LIKES]->(Genre1)
ACTED_IN: Connects actors to movies they acted in.
Example: (Actor1)-[:ACTED_IN]->(Movie1)
DIRECTED: Connects directors to the movies they directed.
Example: (Director1)-[:DIRECTED]->(Movie1)
IN_GENRE: Connects movies to their genres.
Example: (Movie1)-[:IN_GENRE]->(Genre1)

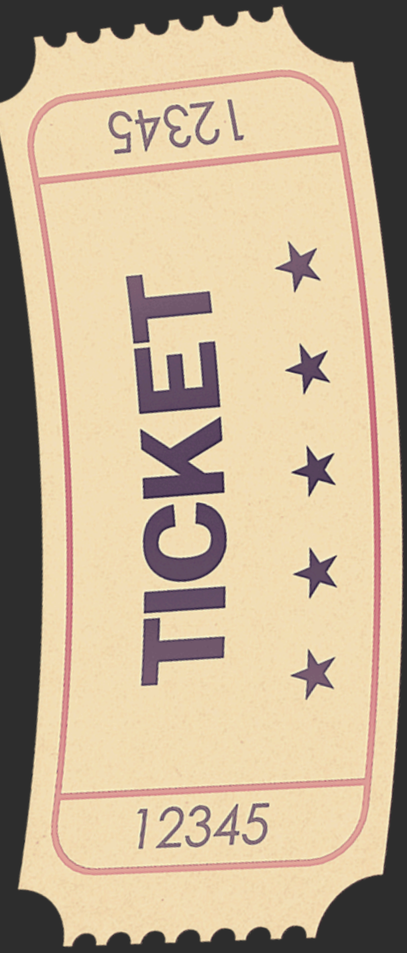
Nodes

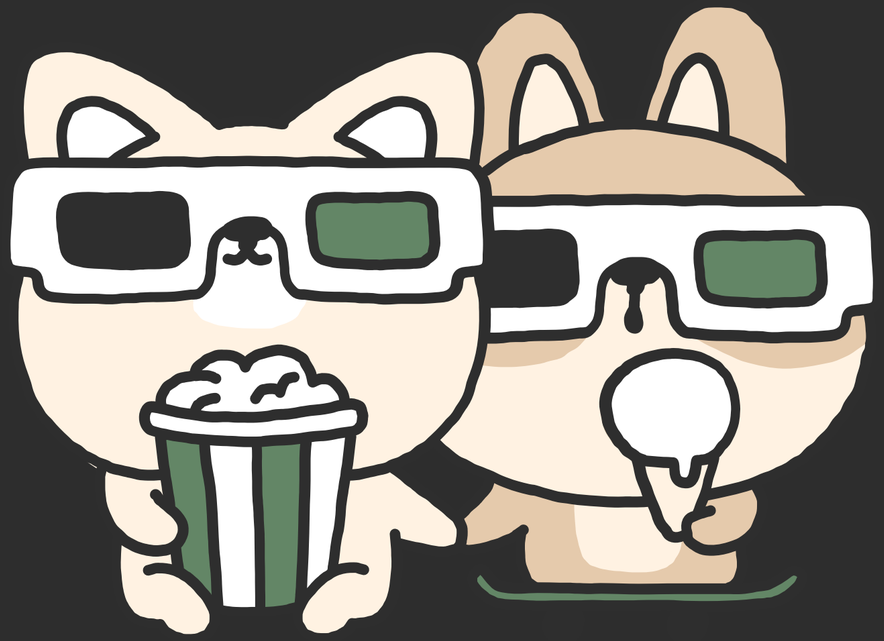
Users: Represent people who interact with movies by providing ratings and showing preferences for genres.
Example: User with attributes like name, surname, birthDate.
Movies: Represent films that users rate or that actors participate in.
Example: Movie with attributes like title, releaseYear.
Actors: Represent individuals who acted in movies.
Example: Actor with attributes like name, surname.
Directors: Represent individuals who directed movies.
Example: Director with attributes like name, surname.
Genres: Represent categories of movies.
Example: Genre with attributes like name.

FINAL VERSION OF A GRAPH









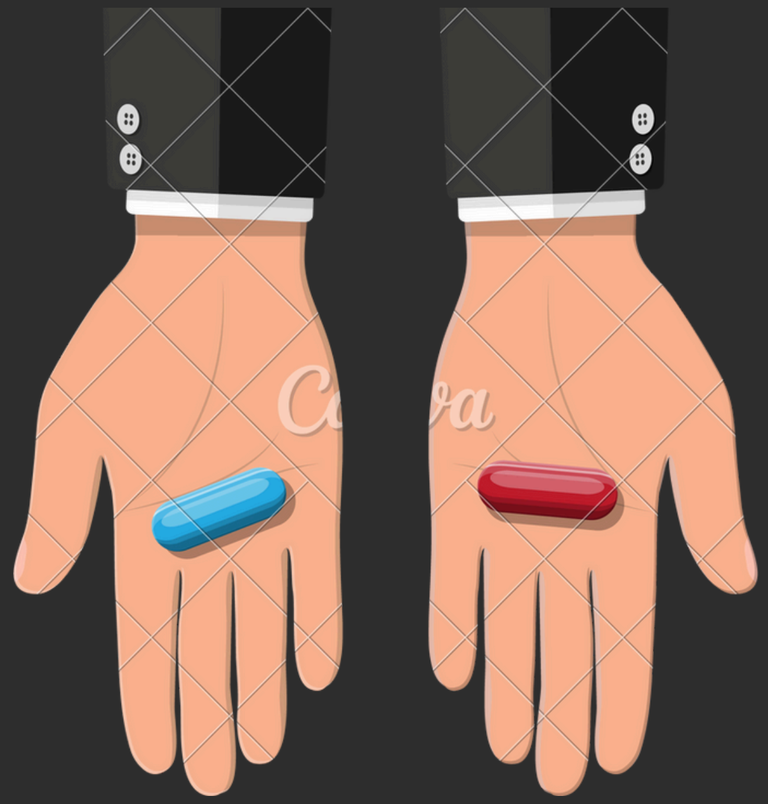
COMPETENCY QUESTIONS

1. What is the average rating for each movie?
2. How many movies did each user rate, and what is their average rating?
3. What movies belong to each genre?
4. Which movie has the most ratings?
5. Which actors have acted in Sci-Fi or Action movies?
6. What is the best-rated movie in each genre?
7. What is the average rating for movies rated by users older than 30?
8. Movie Recommendations Based on Shared Users.

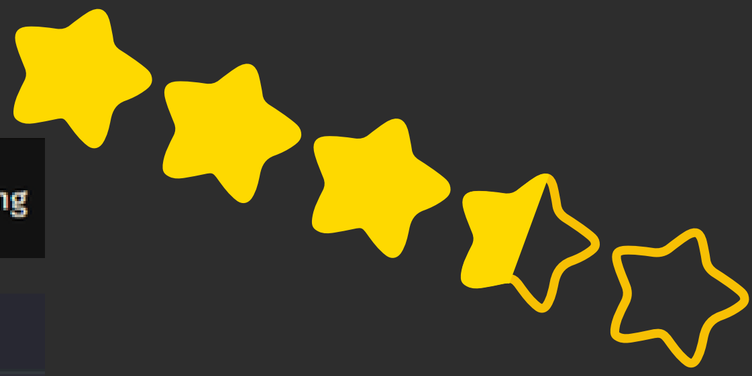
WHAT IS THE AVERAGE RATING FOR EACH MOVIE?

```
MATCH (movie:Movie)←[r:RATED]-(user:User)
RETURN movie.title AS Title, round(AVG(r.rating), 2) AS AverageRating;
```

	Title	AverageRating
1	"Inception"	3.33
2	"Titanic"	3.6
3	"Avatar"	3.0
4	"The Matrix"	4.0

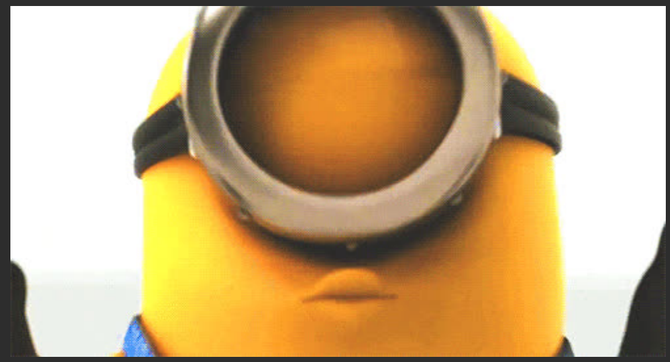


HOW MANY MOVIES DID EACH USER RATE, AND WHAT IS THEIR AVERAGE RATING?



```
MATCH (u:User)-[r:RATED]→(m:Movie)
RETURN u.name AS User, COUNT(r) AS NumberOfRatings, round(AVG(r.rating), 2) AS AverageRating
ORDER BY AverageRating DESC;
```

	User	NumberOfRatings	AverageRating
1	"Charlie"	2	5.0
2	"Eve"	3	4.0
3	"Diana"	2	3.5
4	"Frank"	2	3.5
5	"Bob"	3	3.33
6	"Alice"	3	2.33



WHAT MOVIES BELONG TO EACH GENRE?

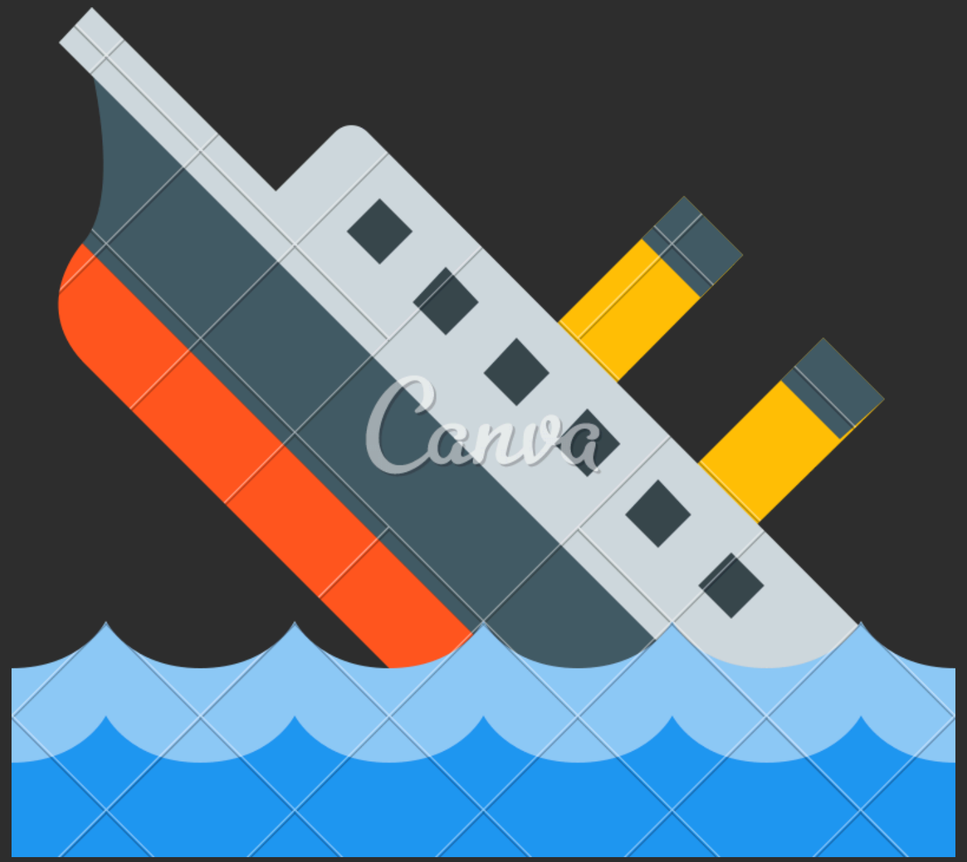
```
MATCH (genre:Genre)←[:IN_GENRE]-(movie:Movie)
RETURN genre.name AS Genre, collect(movie.title) AS Movies;
```

	Genre	Movies
1	"Sci-Fi"	["Inception", "The Matrix", "Avatar"]
2	"Action"	["Inception", "Titanic", "The Matrix"]
3	"Drama"	["Titanic", "Avatar"]

WHICH MOVIE HAS THE MOST RATINGS?

```
MATCH (movie:Movie)←[:RATED]-(user:User)
RETURN movie.title AS Title, COUNT(user) AS NumberOfRatings
ORDER BY NumberOfRatings DESC
LIMIT 1;
```

	Title	NumberOfRatings
1	"Titanic"	5



WHICH ACTORS HAVE ACTED IN SCI-FI OR ACTION MOVIES?

```
MATCH (actor:Actor)-[:ACTED_IN]→(movie:Movie)-[:IN_GENRE]→(genre:Genre)
WHERE genre.name IN ["Sci-Fi", "Action"]
RETURN DISTINCT actor.name AS FirstName, actor.surname AS LastName;
```

	Name	Surname
1	"Leonardo"	"DiCaprio"
2	"Joseph"	"Gordon-Levitt"
3	"Keanu"	"Reeves"
4	"Carrie-Anne"	"Moss"
5	"Laurence"	"Fishburne"
6	"Kate"	"Winslet"
7	"Sam"	"Worthington"
8	"Zoe"	"Saldana"

WHAT IS THE BEST-RATED MOVIE IN EACH GENRE?

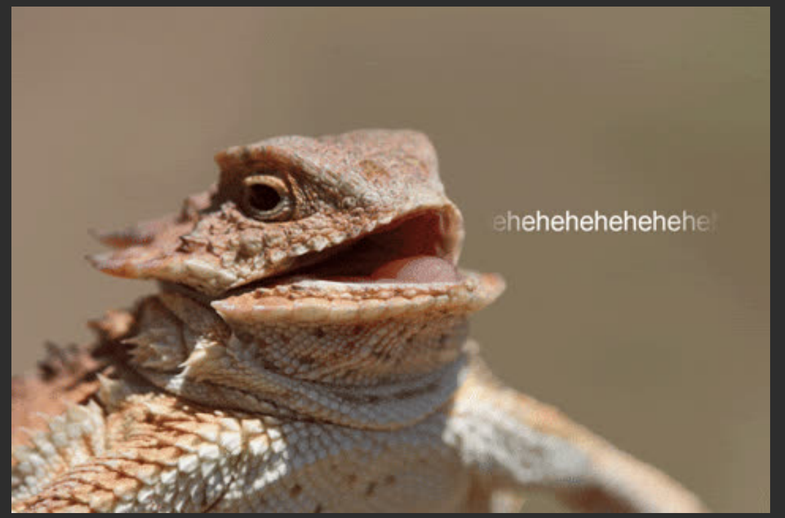
```
MATCH (user:User)-[r:RATED]→(movie:Movie)-[:IN_GENRE]→(genre:Genre)
WITH genre, movie, ROUND(AVG(r.rating),2) AS AverageRating
ORDER BY genre, AverageRating DESC
WITH genre, COLLECT(movie { .title, AverageRating })[0] AS BestMovie
RETURN genre.name AS Genre, BestMovie.title AS Title, BestMovie.AverageRating AS AverageRating
```

	Genre	Title	AverageRating
1	"Sci-Fi"	"The Matrix"	4.0
2	"Drama"	"Titanic"	3.6
3	"Action"	"The Matrix"	4.0

WHAT IS THE AVERAGE RATING FOR MOVIES RATED BY USERS OLDER THAN 30?

```
MATCH (user:User)-[r:RATED]→(movie:Movie)
WHERE duration.between(user.birthDate, date()).years ⩾ 30
RETURN movie.title AS Title, ROUND(AVG(r.rating),2) AS AverageRating, COUNT(r.rating) AS NumberOfRatings
```

	Title	AverageRating	NumberOfRatings
1	"Inception"	4.0	1
2	"Titanic"	4.33	3
3	"Avatar"	2.5	2
4	"The Matrix"	3.5	2



MOVIE RECOMMENDATIONS BASED ON SHARED USERS

```
MATCH (m:Movie {title:'The Matrix'})←[:RATED]-(u:User)-[:RATED]→(rec:Movie)
RETURN
    rec.title AS recommended_movie,
    COUNT(u) AS shared_users,
    COLLECT(u.name) AS shared_user_names
ORDER BY shared_users DESC
```

	recommended_movie	shared_users	shared_user_names
1	"Titanic"	3	["Diana", "Eve", "Frank"]
2	"Inception"	1	["Charlie"]
3	"Avatar"	1	["Eve"]

FUN FACT

THE ICONIC OPENING MOTIF IN MATRIX WAS INSPIRED BY EVERYDAY LIFE

One of the most identifiable images from The Matrix is the green code that runs vertically down the screen. This was designed by the production designer Simon Whitely, he based it on rain running down a window pane, and it’s made up of Japanese katakana characters that Whitely copied from a cookbook.

ANY QUESTIONS?

THANK YOU FOR LISTENING!