Firebase-Based Recipe Analytics Pipeline

Technical Report

Executive Summary

This report documents a complete data engineering pipeline built on Firebase Firestore for recipe data extraction, transformation, validation, and analytics. The system processes an authentic Maharashtrian cuisine dataset containing 20+ recipes, demonstrating modern ETL practices with NoSQL databases.

GitHub Link:- Click here

1. Introduction

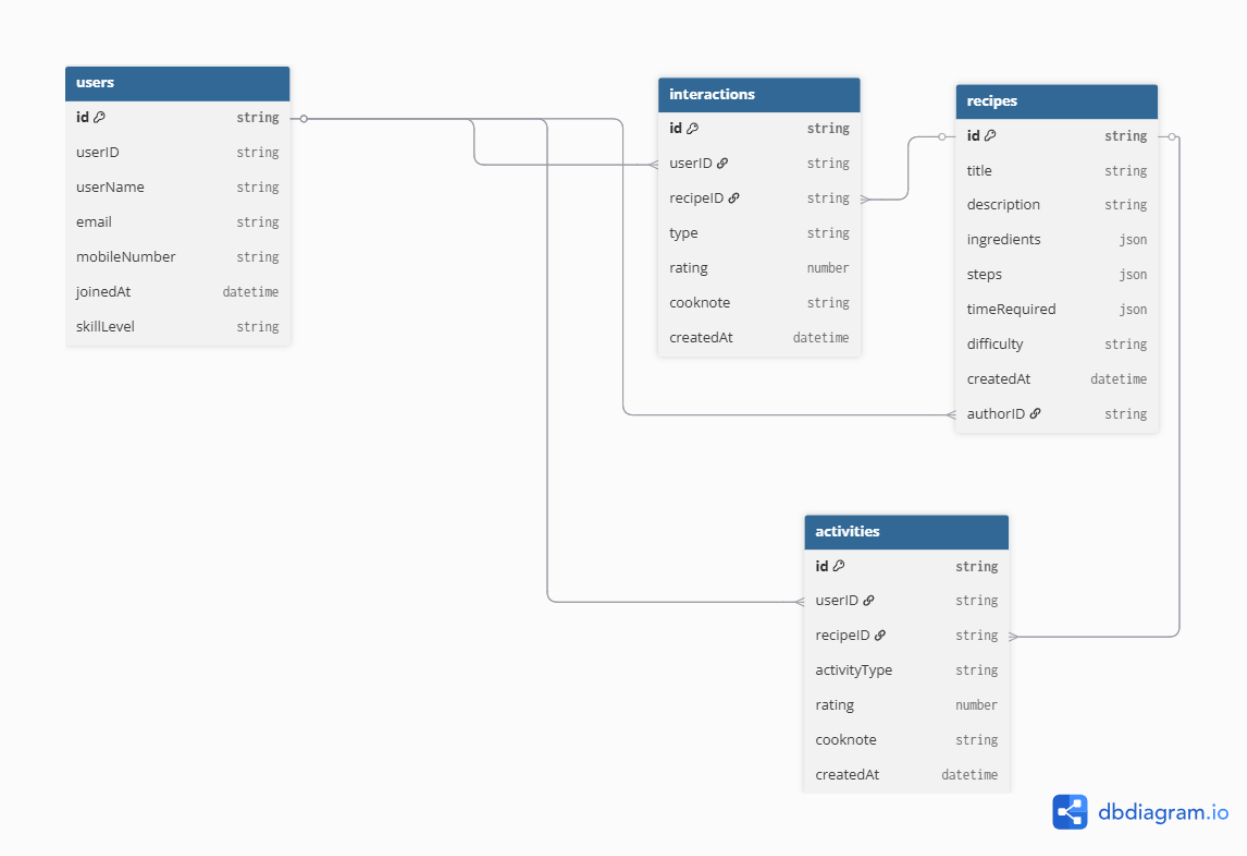
1.1 Project Overview

The Recipe Analytics ETL Pipeline is designed to extract recipe data from Firebase Firestore, transform it into normalized relational formats, validate data quality, and generate actionable insights through analytics and visualizations.

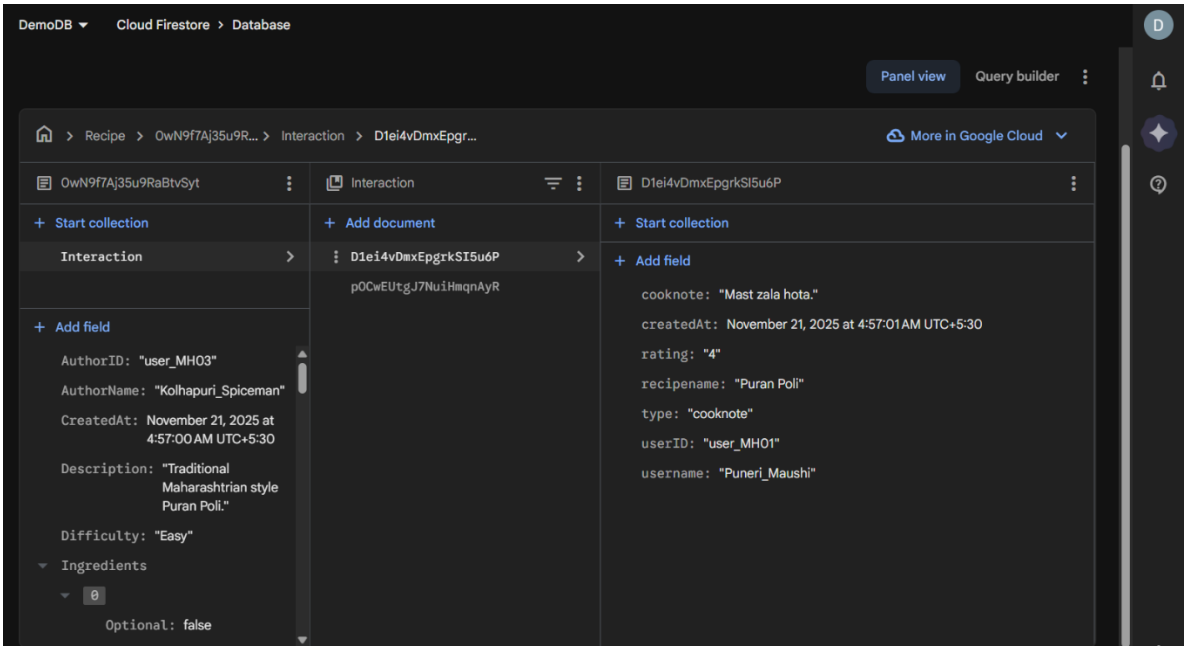
1.2 Technology Stack

Component	Technology
Database	Firebase Firestore
Programming Language	Python 3.8+
Data Processing	Pandas, NumPy
Visualization	Matplotlib
Firebase SDK	firebase-admin

2. Data Model Design



2.1 Database Architecture



The system uses Firebase Firestore with a hierarchical document structure optimized for NoSQL patterns.

2.2 Design Decisions

Design Choice	Reasoning
Interactions under Recipes	Groups recipe activity together; enables fast queries for single recipe analytics
Activities under Users	Tracks user behavior across recipes; enables user-centric analytics
Denormalized author names	Avoids extra reads; Firestore doesn't support JOINs

2.3 Output Schema

The pipeline produces four normalized CSV tables:

recipes.csv - Primary recipe information including title, description, timing, difficulty, category, dietary type, and author details.

	recipe_id	title	description	prep_time_min	cook_time_min	total_time_min	difficulty	category	dietary_type	author_id	author_name	created_at
2	0wN9f7Aj35u9Ra8tvSyt	Puran Poli	Traditional Maharashtrian style Puran Poli.	29	48	77	Easy	Uncategorized	Unknown	user_MH03	Kolhapuri_Spiceman	2025-11-20T23:27:00.8
3	4bdyfBx5qJRHkFXXoCBd	Masale Bhaat	Traditional Maharashtrian style Masale Bhaat.	30	29	59	Hard	Uncategorized	Unknown	user_MH05	Nagpur_Foodie	2025-11-20T23:27:19.1
4	5fh6IZjyhc64yEto3feT	Sabudana Khichdi	Traditional Maharashtrian style Sabudana Khichdi.	34	32	66	Easy	Uncategorized	Unknown	user_MH05	Nagpur_Foodie	2025-11-20T23:26:55.9
5	AG5SNwy0XIIPy2pskV6A	Batata Vada	Traditional Maharashtrian style Batata Vada.	37	25	62	Easy	Uncategorized	Unknown	user_MH02	Mumbai_Cha_Raja	2025-11-20T23:27:22.1
6	ElkULpf1e87utdDBBCh1	Kothimbir Vadi	Traditional Maharashtrian style Kothimbir Vadi.	27	60	87	Medium	Uncategorized	Unknown	user_MH03	Kolhapuri_Spiceman	2025-11-20T23:27:12.6
7	GoZpJ0oPQ8kicn1YLVbZ	Bharli Vangi (Stuffed Eggplant)	Traditional Maharashtrian style Bharli Vangi (Stuffed Eggplant).	30	21	51	Medium	Uncategorized	Unknown	user_MH01	Puneri_Maushi	2025-11-20T23:27:31.8
8	LVmh3o7AD0OJ6viLyXy7	Vada Pav	Traditional Maharashtrian style Vada Pav.	28	55	83	Easy	Uncategorized	Unknown	user_MH04	Konkan_Chef	2025-11-20T23:26:52.1
9	PR60aK9qVaQtGaPTCJmt	Kothimbir Vadi	Traditional Maharashtrian style Kothimbir Vadi.	19	35	54	Medium	Uncategorized	Unknown	user_MH01	Puneri_Maushi	2025-11-20T23:27:25.3
10	VoflyIYaLUSPiRwmITZZ	Kanda Poha	Traditional Maharashtrian style Kanda Poha.	28	29	57	Hard	Uncategorized	Unknown	user_MH02	Mumbai_Cha_Raja	2025-11-20T23:27:27.4
11	Zp0d5EGyhY8q0AVYOPw7	Crispy Dosa	South Indian fermented crepe made from rice and lentil batter.	30	30	540	Medium	Uncategorized	Unknown	user_12345	Dnyaneshwarpotdar	2025-11-20T23:28:22.8
12	aroReC0jFLONYe2tTwol	Kolhapuri Mutton	Traditional Maharashtrian style Kolhapuri Mutton.	21	29	50	Hard	Uncategorized	Unknown	user_MH04	Konkan_Chef	2025-11-20T23:27:35.1
13	bDNmDH9eWFMMEKSN7xjm	Sheera	Traditional Maharashtrian style Sheera.	37	42	79	Easy	Uncategorized	Unknown	user_MH04	Konkan_Chef	2025-11-20T23:27:07.3
14	bZkFBpZeOjehS3oE4mkV	Modak	Traditional Maharashtrian style Modak.	21	33	54	Hard	Uncategorized	Unknown	user_MH04	Konkan_Chef	2025-11-20T23:27:38.8
15	fKFhRLiD7fbesBzQDZeu	Thalipeeth	Traditional Maharashtrian style Thalipeeth.	27	28	55	Easy	Uncategorized	Unknown	user_MH05	Nagpur_Foodie	2025-11-20T23:27:45.3
16	l8UGZEWw0ijk1d5zoNfM	Shrikhand	Traditional Maharashtrian style Shrikhand.	20	60	80	Hard	Uncategorized	Unknown	user_MH01	Puneri_Maushi	2025-11-20T23:27:02.8

ingredients.csv - Normalized ingredient data linked to recipes via foreign key, containing name, quantity, unit, and optional flag.

1	recipe_id	name	quantity	unit	is_optional
2	0wN9f7Aj35u9RaBtvSyt	Ghee	131	vati (bowl)	False
3	0wN9f7Aj35u9RaBtvSyt	Bombay Duck (Fish)	25	tsp	False
4	0wN9f7Aj35u9RaBtvSyt	Jaggery (Gul)	20	cups	True
5	0wN9f7Aj35u9RaBtvSyt	Goda Masala	199	vati (bowl)	False
6	0wN9f7Aj35u9RaBtvSyt	Sabudana	107	vati (bowl)	False
7	0wN9f7Aj35u9RaBtvSyt	Curry Leaves	62	tbsp	False
8	0wN9f7Aj35u9RaBtvSyt	Oil	196	tbsp	True
9	0wN9f7Aj35u9RaBtvSyt	Rice Flour	200	tsp	False
10	4bdyFBx5qJRHkFXXoCBd	Ginger	111	tsp	False
11	4bdyFBx5qJRHkFXXoCBd	Jaggery (Gul)	187	vati (bowl)	False
12	4bdyFBx5qJRHkFXXoCBd	Mustard Seeds	39	cups	False
13	4bdyFBx5qJRHkFXXoCBd	Green Chilies	231	pieces	False
14	4bdyFBx5qJRHkFXXoCBd	Tamarind	217	tbsp	False
15	4bdyFBx5qJRHkFXXoCBd	Chana Dal	68	grams	False
16	4bdyFBx5qJRHkFXXoCBd	Poha (Flattened Rice)	86	vati (bowl)	True
17	4bdyFBx5qJRHkFXXoCBd	Goda Masala	177	cups	False
18	4bdyFBx5qJRHkFXXoCBd	Matki (Moth Beans)	104	tsp	False
19	4bdyFBx5qJRHkFXXoCBd	Rice Flour	5	tsp	True
20	4bdyFBx5qJRHkFXXoCBd	Grated Coconut	11	tbsp	False
21	4bdyFBx5qJRHkFXXoCBd	Besan (Gram Flour)	55	tbsp	False
22	5fh6lZjyhc64yEto3feT	Coriander Leaves	68	grams	True
23	5fh6lZjyhc64yEto3feT	Green Chilies	101	vati (bowl)	True

steps.csv - Recipe instructions with step numbers, instructions, and duration, linked to parent recipes.

	recipe_id	step_number	instruction	duration
2	0wN9f7Aj35u9RaBtvSyt	1	Sauté the mixture properly for authentic taste.	9 min
3	0wN9f7Aj35u9RaBtvSyt	2	Deep Fry the mixture properly for authentic taste.	7 min
4	0wN9f7Aj35u9RaBtvSyt	3	Tempering (Phodni) the mixture properly for authentic taste.	9 min
5	0wN9f7Aj35u9RaBtvSyt	4	Garnish the mixture properly for authentic taste.	4 min
6	0wN9f7Aj35u9RaBtvSyt	5	Pressure Cook the mixture properly for authentic taste.	9 min
7	4bdyFBx5qJRHkFXXoCBd	1	Steam the mixture properly for authentic taste.	12 min
8	4bdyFBx5qJRHkFXXoCBd	2	Garnish the mixture properly for authentic taste.	8 min
9	4bdyFBx5qJRHkFXXoCBd	3	Sauté the mixture properly for authentic taste.	13 min
10	4bdyFBx5qJRHkFXXoCBd	4	Steam the mixture properly for authentic taste.	9 min
11	4bdyFBx5qJRHkFXXoCBd	5	Sauté the mixture properly for authentic taste.	11 min
12	4bdyFBx5qJRHkFXXoCBd	6	Steam the mixture properly for authentic taste.	5 min
13	4bdyFBx5qJRHkFXXoCBd	7	Steam the mixture properly for authentic taste.	5 min
14	4bdyFBx5qJRHkFXXoCBd	8	Deep Fry the mixture properly for authentic taste.	4 min
15	4bdyFBx5qJRHkFXXoCBd	9	Roll the mixture properly for authentic taste.	14 min
16	5fh6lZjyhc64yEto3feT	1	Deep Fry the mixture properly for authentic taste.	8 min
17	5fh6lZjyhc64yEto3feT	2	Knead the mixture properly for authentic taste.	6 min
18	5fh6lZjyhc64yEto3feT	3	Steam the mixture properly for authentic taste.	5 min
19	5fh6lZjyhc64yEto3feT	4	Pressure Cook the mixture properly for authentic taste.	4 min
20	AG5SNwy0XIIPy2pskV6A	1	Steam the mixture properly for authentic taste.	6 min
21	AG5SNwy0XIIPy2pskV6A	2	Deep Fry the mixture properly for authentic taste.	14 min
22	AG5SNwy0XIIPy2pskV6A	3	Knead the mixture properly for authentic taste.	12 min
23	AG5SNwy0XIIPy2pskV6A	4	Tempering (Phodni) the mixture properly for authentic taste.	11 min

interactions.csv - User engagement data including ratings, cook notes, and timestamps.

	interaction_id	recipe_id	user_id	username	type	rating	cooknote	created_at
2	D1ei4vDmxEpgrkSI5u6P	0wN9f7Aj35u9RaBtvSyt	user_MH01	Puneri_Maushi	cooknote	4	Mast zala hota.	2025-11-20T23:27:01.223200+00:00
3	pOCwEUtgJ7NuiHmqnAyR	0wN9f7Aj35u9RaBtvSyt	user_MH05	Nagpur_Foodie	like	4	Mast zala hota.	2025-11-20T23:27:02.494744+00:00
4	iNW8dAYZI6gqQW0X19DC	4bdyFBx5qJRHkFXXoCBd	user_MH02	Mumbai_Cha_Raja	cooknote	4	Mast zala hota.	2025-11-20T23:27:20.849194+00:00
5	pm6G5T5UBVKk2MzYrDGR	4bdyFBx5qJRHkFXXoCBd	user_MH01	Puneri_Maushi	rating	5	Lai bhari! (Awesome)	2025-11-20T23:27:20.513229+00:00
6	kzDhZIEF6PMe6icZxhUG	5fh6lZjyhc64yEto3feT	user_MH01	Puneri_Maushi	share	4	Lai bhari! (Awesome)	2025-11-20T23:26:57.934710+00:00
7	mDv35UAUZn4HjlSGGSMm	5fh6lZjyhc64yEto3feT	user_MH04	Konkan_Chef	like	4	Mast zala hota.	2025-11-20T23:26:57.557348+00:00
8	vqJkNyhKfGToRKQN7dsn	5fh6lZjyhc64yEto3feT	user_MH05	Nagpur_Foodie	view	4	Lai bhari! (Awesome)	2025-11-20T23:26:56.306906+00:00
9	DM4J9LKXg58AOCy8PmeC	AG5SNwy0XIIPy2pskV6A	user_MH02	Mumbai_Cha_Raja	like	3	Mast zala hota.	2025-11-20T23:27:24.215641+00:00
10	Pv54mNCWKlcigq4I8F1V	AG5SNwy0XIIPy2pskV6A	user_MH03	Kolhapuri_Spiceman	view	3	Lai bhari! (Awesome)	2025-11-20T23:27:22.584832+00:00
11	hFiWN4qulrirew3eN4Jx	AG5SNwy0XIIPy2pskV6A	user_MH05	Nagpur_Foodie	share	3	Mast zala hota.	2025-11-20T23:27:23.753995+00:00
12	8P3lafx26dE8jDvU9CZD	ElkULpf1e87utdDBBCh1	user_MH04	Konkan_Chef	rating	3	Lai bhari! (Awesome)	2025-11-20T23:27:13.951325+00:00
13	II9EtTXrsncCBUWNK6wq	ElkULpf1e87utdDBBCh1	user_MH02	Mumbai_Cha_Raja	like	4	Mast zala hota.	2025-11-20T23:27:14.273621+00:00
14	3XpjkOxVDmjg7pRXlp0T	GoZpJ0oPQ8kicn1YLVbZ	user_MH05	Nagpur_Foodie	like	3	Mast zala hota.	2025-11-20T23:27:33.480395+00:00
15	V0a2YK3mY7FOKHqznb7A	GoZpJ0oPQ8kicn1YLVbZ	user_MH01	Puneri_Maushi	share	5	Mast zala hota.	2025-11-20T23:27:32.316629+00:00
16	ef9qunDQq43ZFXwMT6DM	GoZpJ0oPQ8kicn1YLVbZ	user_MH05	Nagpur_Foodie	share	3	Lai bhari! (Awesome)	2025-11-20T23:27:33.953655+00:00
17	XOi3cKbgfdMyNV7Gyn3w	LVmh3o7AD0OJ6viLyXy7	user_MH02	Mumbai_Cha_Raja	view	4	Mast zala hota.	2025-11-20T23:26:53.012154+00:00
18	M17kG8ZVG98HGNbmqLSZ	PR60aK9qVaQtGaPTCJmt	user_MH02	Mumbai_Cha_Raja	like	4	Mast zala hota.	2025-11-20T23:27:25.834002+00:00
19	t9zPmQb3M35QbvArdrjp	PR60aK9qVaQtGaPTCJmt	user_MH05	Nagpur_Foodie	rating	4	Mast zala hota.	2025-11-20T23:27:27.026199+00:00
20	1CzRwEgDs21mEAwOPX3D	VoflylYaLUSPiRwmITZZ	user_MH01	Puneri_Maushi	share	5	Lai bhari! (Awesome)	2025-11-20T23:27:28.665342+00:00
21	GH1IgQmnONRoyVkcVF1P	VoflylYaLUSPiRwmITZZ	user_MH01	Puneri_Maushi	rating	3	Lai bhari! (Awesome)	2025-11-20T23:27:30.263044+00:00
22	Z9k4UkIESChBXW8hUnrG	VoflylYaLUSPiRwmITZZ	user_MH04	Konkan_Chef	cooknote	3	Lai bhari! (Awesome)	2025-11-20T23:27:29.071681+00:00

---

3. ETL Process

3.1 Pipeline Architecture

The pipeline follows a four-stage process: Extract → Transform → Validate → Analyze

3.2 Extract Phase

The extraction phase connects to Firebase Firestore using the Admin SDK with streaming for efficient data retrieval. Key operations include authentication via service account credentials, document streaming for large collections, subcollection extraction for interactions, and timestamp conversion to ISO 8601 format.

3.3 Transform Phase

Transformation	Description
Flatten Ingredients	Nested array converted to separate CSV with recipe_id FK
Flatten Steps	Nested array converted to separate CSV with recipe_id FK
Extract Subcollections	Firestore subcollection to interactions.csv
Normalize Time	Structured time object to individual minute columns
Handle Missing Data	Default values: "Uncategorized", "Unknown"

---

4. Data Validation

4.1 Validation Rules

Rule	Field	Criteria
Required Fields	title	Must not be empty
Valid Difficulty	difficulty	Must be: Easy, Medium, Hard, Expert
Prep Time	prep_time_min	Must be > 0
Cook Time	cook_time_min	Must be ≥ 0
Time Logic	total_time_min	Must be ≥ prep_time + cook_time
Ingredient Quantity	quantity	Must be > 0 if numeric
Rating Range	rating	Must be between 0 and 5
Has Steps	steps	At least one step required
Has Ingredients	ingredients	At least one ingredient required

4.2 Validation Output

The validator produces a JSON report containing total recipe count, valid/invalid counts, detailed error messages for invalid records, and list of valid record IDs.

---

5. Analytics & Insights

5.1 Generated Insights

The pipeline produces 11 analytical insights:

1. **Most Common Ingredients** - Top 20 ingredients by frequency
2. **Average Prep Time** - Mean preparation time in minutes
3. **Average Cook Time** - Mean cooking time in minutes
4. **Difficulty Distribution** - Recipe count per difficulty level
5. **Most Interacted Recipes** - Top 20 by interaction count
6. **Prep vs Rating Correlation** - Statistical correlation analysis
7. **High-Rating Ingredients** - Ingredients appearing in 4+ star recipes
8. **Top Rated Recipes** - Top 10 by average rating
9. **Steps Distribution** - Statistical summary of recipe complexity
10. **Most Commented Recipes** - Top 10 by cook note count
11. **Longest Recipes** - Top 10 by total preparation time

5.2 Visualizations

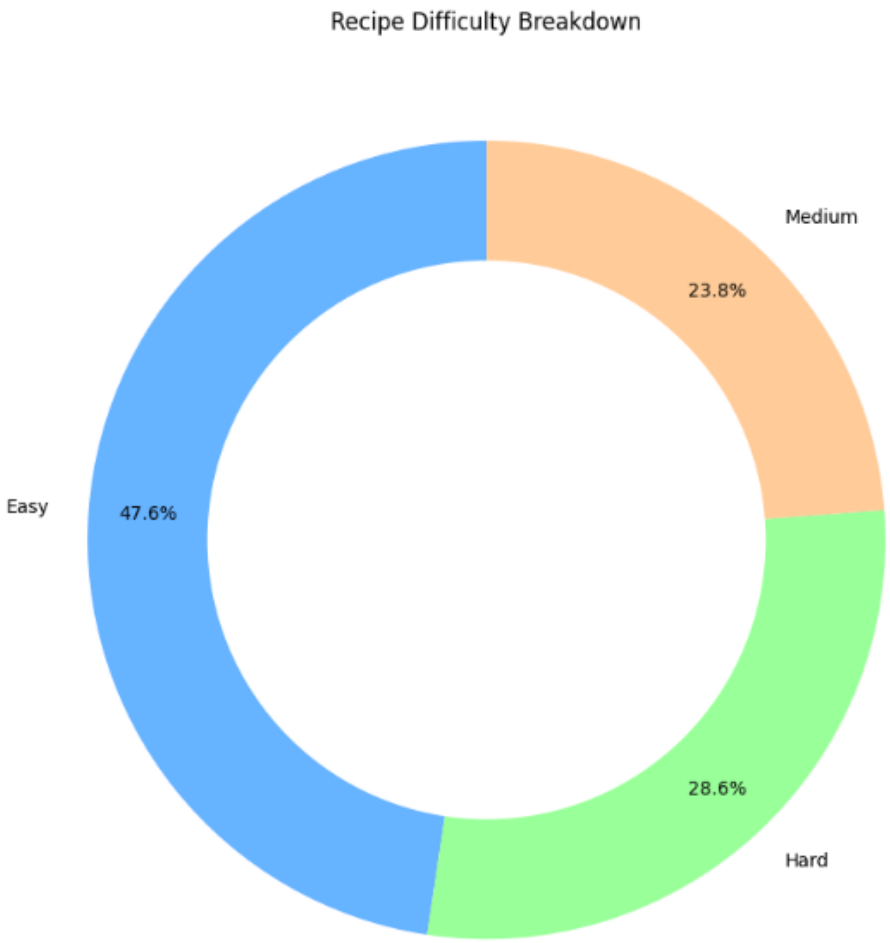


Fig 1.1 Donut Chart

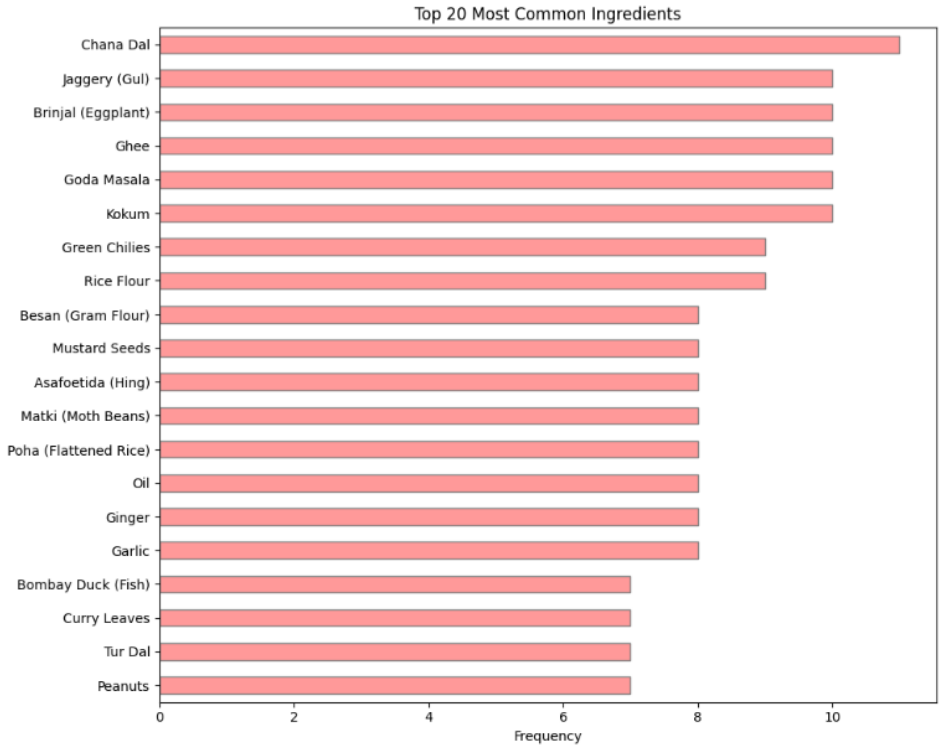


Fig 1.2 Horizontal Bar

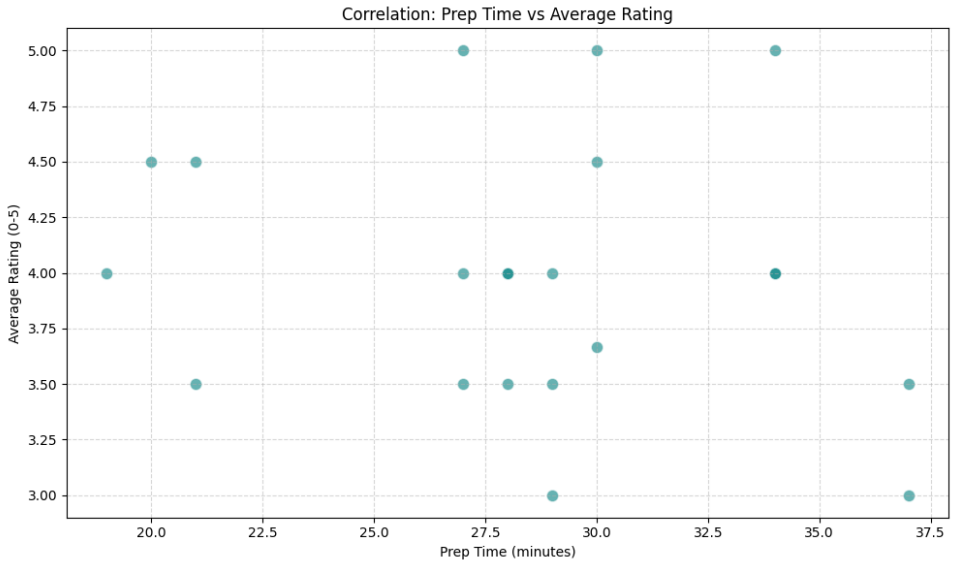


Fig 1.3 Scatter Plot

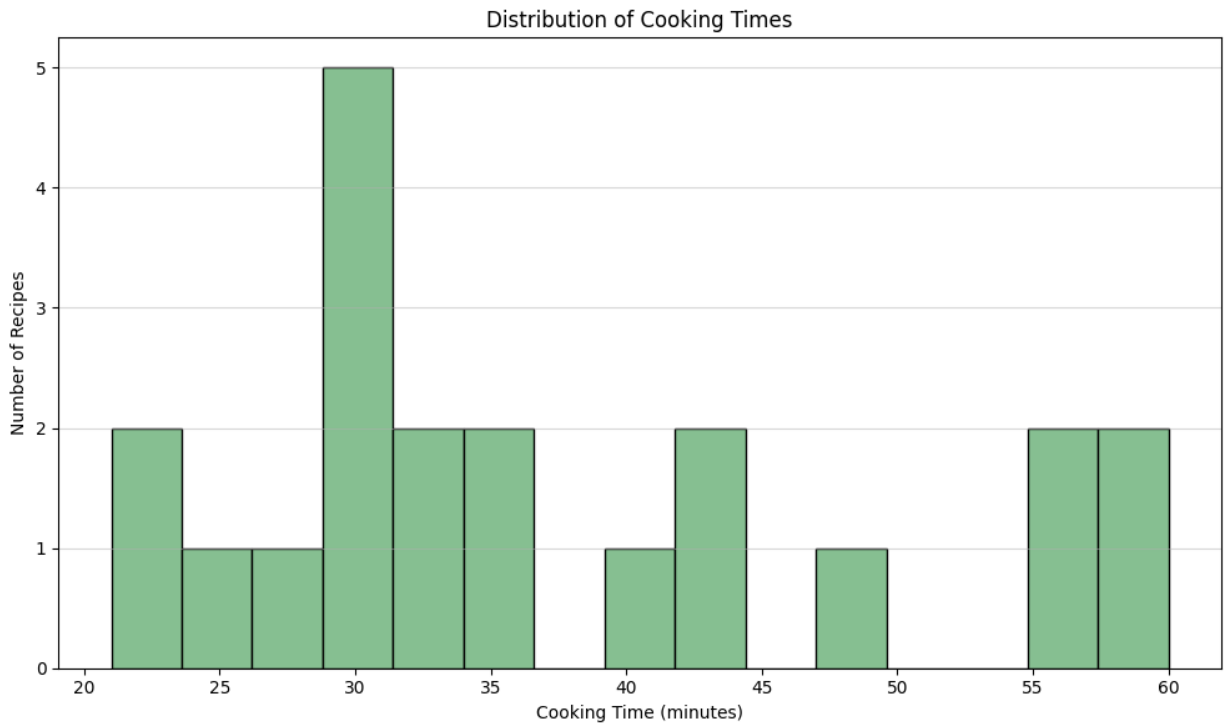


Fig 1.4 Horizontal Bar

Chart	**Type**	**Description**
difficulty_donut_chart.png	Donut Chart	Recipe distribution by difficulty
top_ingredients_bar_chart.png	Horizontal Bar	Top 20 most used ingredients
prep_vs_rating_scatter_plot.png	Scatter Plot	Correlation between prep time and rating
cook_time_histogram.png	Histogram	Distribution of cooking times

6. Constraints & Limitations

6.1 Firestore Limitations

Constraint	Impact	Mitigation
No native JOINs	Cannot query across collections	Denormalized data; subcollections
Read costs	Each document read is billed	Streaming instead of batch reads
No aggregations	No COUNT/SUM/AVG in queries	Aggregations in Python post-export
Subcollection queries	Cannot query all at once	Iterate per parent document

6.2 Pipeline Constraints

- Sequential execution required (seed → transform → validate → analyze)
- Full export only; no incremental/delta processing
- Memory-bound with pandas DataFrames
- Hardcoded relative paths in some scripts

6.3 Scalability Notes

- Current tested capacity: ~20 recipes, ~200 interactions
- For 1000+ recipes: implement pagination in export
- For larger datasets: consider chunked processing or Apache Spark

---

7. Project Structure

```
RECIPE_ANALYTICS/
├── analytics/          # Analytics outputs and charts
├── config/             # Firebase credentials
├── data_validation/    # Validation scripts and reports
├── Firebase_Setup/     # Data seeding scripts
├── transform_data/     # ETL outputs (CSV files)
└── README.md
```

---

8. Installation & Execution

8.1 Prerequisites

- Python 3.8 or higher
- Firebase project with Firestore enabled
- Service account credentials (JSON)

8.2 Execution Steps

1. **Seed Initial Data** - Run seed_data.py to create base recipe and user
2. **Generate Synthetic Data** - Run genrate_sytetic.py for 20 Maharashtrian recipes
3. **Transform Data** - Run transform.py to export to CSV
4. **Validate Data** - Run validator.py for quality checks
5. **Generate Analytics** - Run analytics.py for insights and charts

---

9. Deliverables Summary

Deliverable	Status
Source files for ETL scripts	Complete
Validation script	Complete
Normalized CSV output	Complete
Analytics summary (JSON)	Complete
Documentation	Complete
Visualization charts	Complete

---

10. Conclusion

The Firebase-Based Recipe Analytics Pipeline successfully demonstrates a complete ETL workflow for NoSQL data, producing normalized relational outputs suitable for further analysis. The system handles the unique challenges of Firestore's document model while maintaining data quality through comprehensive validation.

Author: Dnyaneshwar Potdar