QuickBite Data Pipeline – README

# 📌 Project Overview

This project demonstrates an end-to-end data pipeline using AWS services for QuickBite, a food delivery platform. It processes order data, applies data cleaning and enrichment, and loads it into Amazon Redshift for analytics.

# 📁 S3 Structure

Bucket: s3://quickbite-data-suraj/  
├── Landing/  
│ └── sample\_orders.csv  
├── Trusted/  
│ └── transactions/ (Partitioned Parquet Output)  
│ └── dim\_customer.csv

# 🔧 AWS Glue ETL

- Glue script written in PySpark  
- Cleaned data: null filtering, casting, calculated fields  
- Partitioned by year and month

# 🔍 Athena

- Crawler created for Trusted/transactions  
- Queries run on `transactions` table to explore and filter data

# 📊 Redshift

- Redshift cluster created and connected via Query Editor v2  
- Used COPY command to load trusted data  
- Created tables: `fact\_order`, `dim\_customer`  
- Performed joins for customer analytics

# 📈 Data Quality Checks

See `Data\_Quality\_Report\_QuickBite.docx` for:  
- Null analysis  
- Duplicate checks  
- Order status counts  
- Fixes applied

# ✅ Completed On

2025-06-10