

**ANALYSIS OF ENERGY CONSUMPTION
PATTERNS IN THE UNITED STATES**
Version 1.1

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

Revision History

Date	Version	Description	Author
03/10/2023	1.0	Version 1 Project topic finalization, Dataset selection, Project deliverables	Manidedeepya Chennapragada, Kalyan Vikkurthi, Dharnidhar Reddy Banala, Nikhil Mylarusetty
05/05/2023	1.1	Version 1.1 Design, Implementation, and documentation	Manidedeepya Chennapragada, Kalyan Vikkurthi, Dharnidhar Reddy Banala, Nikhil Mylarusetty

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

Table of Contents

1.	Introduction	4
	1.1 Purpose of this document	4
	1.2 Intended Audience	4
	1.3 Scope	4
	1.4 Definitions and acronyms	4
	1.4.1 Definitions	4
	1.4.2 Acronyms and abbreviations	5
	1.5 References	5
2.	Background and Objectives	5
3.	Organization	6
	3.1 Project group	6
	3.2 Customer	7
4.	Development process	7
5.	Deliverables	7
6.	Project risks	8
7.	Communication	8
	7.1 Canvas	Error! Bookmark not defined.
	7.2 Git	8
8.	Project plan	8
	8.1 Time schedule	8
	8.1.1 Remarks	9
	8.2 Test plan	11
	8.2.1 Testing Remarks	12
9.	References	12

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

1. Introduction

1.1 Purpose of this document

The purpose of this document is to provide a detailed project description of the application called Analysis Of Energy Consumption Patterns In The United States, which is designed to help people understand the electricity consumption trends. This document includes details about organization, roles, deliverables, project risks, time plans and financial plans.

1.2 Intended Audience

This document shall be used in all phases of the project as a guideline. Intended audiences of this project are all project stakeholders:

- Professor Andrew H. Bond
- Kalyan Vikkurthi
- Manideepya Chennapragada, Dharnidhar Reddy Banala, Nikhil Mylarusetty

1.3 Scope

This document defines the project plan of the Analyzing consumer shopping behavior from a large multi-category online store application. The overview includes objectives of the project, organization of the project team, development process that is going to be used during the project, assessment of possible risks, communication used between project stakeholders and project plan that includes time schedule and activity plan.

1.4 Definitions and acronyms

1.4.1 Definitions

Keyword	Definitions
Project Name	Analysis Of Energy Consumption Patterns In The United States
Project Supervisor	Professor Andrew H. Bond
Project Leader	Kalyan Vikkurthi
Team Member	Manideepya Chennapragada, Dharnidhar Reddy Banala, Nikhil Mylarusetty
Milestone	05/07/2023
Git	https://github.com/KalyanVikkurthi002/Datawizards-kafka
Scrum	An iterative and incremental agile software development method for managing software projects and product or application development
Kunagi	Web-based tool for integrated agile project management and collaboration based on Scrum
Scrum sprint	The basic unit of development in Scrum
Scrum master	Dharnidhar Reddy Banala
Product owner	Kalyan Vikkurthi

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

1.4.2 Acronyms and abbreviations

Acronym or abbreviation	Definitions
AWS	Amazon Web Services
ELT	Extract, Load, Transform
IAM	Identity & Access Management
EDA	Exploratory Data Analysis
CLI	Command Line Interface

1.5 References

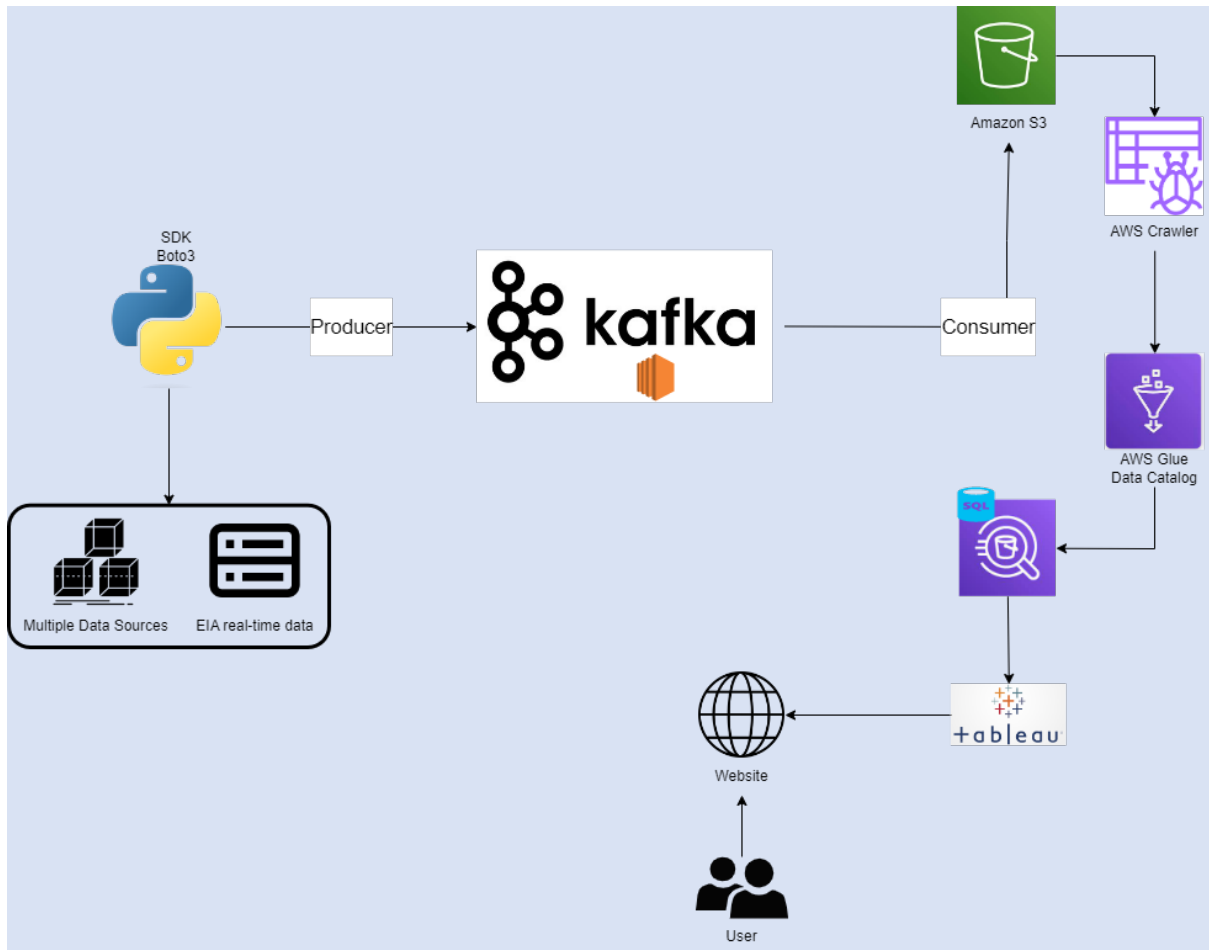
1. <https://www.eia.gov/opendata/browser/electricity/electric-power-operational-data?frequency=monthly&data=ash-content;consumption-for-eg;consumption-for-eg-btu;consumption-uto;consumption-uto-btu;cost;cost-per-btu;generation;heat-content;receipts;receipts-btu;stocks;sulfur-content;total-consumption;total-consumption-btu;&start=2020-01&end=2022-12&sortColumn=period;&sortDirection=desc;>
- 2.
3. <http://www.scrum.org/>
4. <http://kunagi.org/>

2. Background and Objectives

Our project aims to analyze energy consumption patterns in the United States using big data technologies. To achieve this, we are leveraging various technologies, including Apache Kafka, Amazon S3, AWS Crawler, AWS Glue, AWS Athena, Jupyter Notebook, and Tableau. The project builds on the work of the US Energy Information Administration (EIA), a government agency responsible for collecting, analyzing, and disseminating energy-related data and information. The EIA provides independent and impartial energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment.

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

3. Architecture & High-Level Design



4. Organization

4.1 Project group

Name	Initials	Responsibility (roles)
Manidedeepya Chennapragada	MC	Data Transformation, Athena Query, Documentation
Dharnidhar Reddy Banala	DB	Data Visualization, Web Integration, Documentation
Kalyan Vikkurthi	KV	Data Loading, Data Transformation, Documentation
Nikhil Mylarusetty	NM	Web Integration, Athena Query, Documentation

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

4.2 Customer

The target customers are listed below:

- Policymakers
- Researchers
- Energy Companies

5. Development process

The data processing for this project involved importing data from the EIA using the AWS CLI for API data upload. The team performed data cleaning and transformation, which included removing duplicates, handling missing values, and converting data types. Real-time data was obtained and streamed using Kafka, with data stored in an AWS S3 bucket and crawled using AWS Crawler for further processing. A schema was generated in AWS Glue to define the structure and organization of the data, and AWS Athena enabled querying of the data. Tableau was connected to Amazon Athena for data visualization. An authentication-enabled login page was developed for authorized users to access the website, which included an electricity analysis user dashboard. These processes ensured that the data was accurate, consistent, and suitable for subsequent processing and insights extraction.

6. Deliverables

To	Output	Planned week	Promised week	Late +/-	Delivered week	Notes
	Gathering data from EIA website	Feb Third Week	Feb Third Week		02/18/2023	
	Kafka Connection	March Second Week	March Second Week		03/10/2023	
	Loaded the data into S3	March Last Week	March Last Week		03/25/2023	
	Transformed the data	April First Week	April First Week		04/04/2023	
	Querying Via Athena	April Second Week	April Second Week	+1	04/12/2023	
	Data Visualizations	April Third Week	April Third Week	+1	04/20/2023	
	Static Website	April Last Week	April Last Week	+1	04/28/2023	
	Project Report & PPT	May First Week	May First Week		05/01/2023	

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

7. Project risks

Possibility	Risk	Preventive action
Cost risk for using AWS services (15GB Data)	High Cost	Deleted the files and resources not in use.

8. Communication

Had weekly meetings over zoom call for project discussions and implementation.

8.1 Collaboration

8.2 Git

All source code and finished documentation will be uploaded to Github repository.

Repository URL: <https://github.com/KalyanVikkurthi002/Datawizards-kafka>

9. Project plan

9.1 Time schedule

Id	Milestone Description	Responsible Dept./Initials	Finished week Plan	Forecast Week	+/-	Actual	Metr.	Rem.
1	Gathering data from EIA website	MC	Feb Third Week			02/18/2023		
		KV	March Second Week			03/10/2023		
2	Loaded the data into S3	DB , NM	March Last Week			03/25/2023		
3	Transformed the data	KV, MC	April First Week			04/04/2023		
4	Querying Via Athena	SS, IB	April Second Week		+1	04/12/2023		
5	Data Visualizations	DB	April Third Week		+1	04/20/2023		

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

6	Website	NM,MC	April Last Week		+1	04/28/2023			
---	---------	-------	-----------------------	--	----	------------	--	--	--

9.1.1 Remarks

Remark Id	Description

9.2 Test plan

Test No.	001	Phase:	1	Author:	Manidedeepya Chennapragada	Date: 03/03/2023
Test Category:	System Test					
Software Product:	Python					
Test Title:	Pulling data from EIA website using RESTful APIs					
Test Purpose:	To extract data from the EIA website using RESTful APIs					
Test Setup:	Jupyter Notebook					
Prerequisites:	Internet connection, EIA API credentials					
Procedure:	Utilize appropriate RESTful APIs to extract data from the EIA website					
Checks:	Verify the retrieved data in the Jupyter Notebook for completeness and accuracy					
Expected Results:	Able to load the complete data into the Jupyter notebook without any data leakage as per the provided parameters					
Result:	Able to load the complete data into the Jupyter notebook without any data leakage as per the provided parameters					
Reason for Failure:	-					
Remarks:	Successfully completed					

Test No.	002	Phase:	2	Author:	Kalyan Vikkurthi	Date: 03/18/2023
Test Category:	System Test					

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

Software Product:	Apache Kafka
Test Title:	Sending data to Kafka cluster with the help of producer
Test Purpose:	To test the ability to send obtained data to the Kafka cluster using a producer
Test Setup:	Jupyter Notebook
Prerequisites:	Kafka cluster, producer component setup
Procedure:	Use the producer component to send the obtained data to the Kafka cluster
Checks:	Verify the successful transfer of complete data from the producer to the consumer component
Expected Results:	To receive complete data sent by the producer to the consumer component without any data loss
Result:	Able to receive complete data sent by the producer to the consumer component without any data loss
Reason for Failure:	-
Remarks:	Successfully completed

Test No.	003	Phase:	3	Author:	Nikhil Mylarusetty	Date: 04/04/2023
Test Category:	System Test					
Software Product:	S3, Glue, Athena					
Test Title:	Storing results into S3 bucket					
Test Purpose:	To test the ability to store the obtained results into an S3 bucket					
Test Setup:	Jupyter Notebook					
Prerequisites:	AWS S3 setup, consumer component setup					
Procedure:	Utilize the consumer component to write the obtained data into the designated S3 bucket					
Checks:	Verify the successful storage of the data by checking if it is correctly written into the S3 bucket					
Expected Results:	Consumer component should successfully write the data into the S3 bucket					
Result:	Consumer component is able successfully write the data into the S3 bucket					
Reason for Failure:	-					

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

Remarks:	Successfully completed
----------	------------------------

Test No.	004	Phase:	4	Author:	Dharanidhar Reddy Banala	Date: 04/18/2023
Test Category:	System Test					
Software Product:	Tableau Public Server					
Test Title:	Querying results in AWS Athena					
Test Purpose:	To test the querying capabilities in AWS Athena					
Test Setup:	AWS Athena, S3					
Prerequisites:	AWS S3 bucket, crawler job, schema setup					
Procedure:	Create a crawler job in AWS to read the data from the S3 bucket and generate a schema in AWS Data Catalog					
Checks:	Check if the crawler job successfully reads the data and creates a schema in AWS Data Catalog					
Expected Results:	Crawler job should successfully read the data from the S3 bucket and create a successful schema in AWS Data Catalog					
Result:	-					
Reason for Failure:	-					
Remarks:	Successfully completed					

Test No.	005	Phase:	5	Author:	Nikhil Mylarusetty, Manidedeepya Chennapragada	Date: 05/01/2023
Test Category:	System Test					
Software Product:	Angular					
Test Title:	Viewing data statistics in deployed website					
Test Purpose:	To test the ability to view data statistics in the deployed website					
Test Setup:	Deployed website					
Prerequisites:	Authorized user login credentials					
Procedure:	Login to the deployed website as each team member and verify if they can see their user dashboard					

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023

Checks:	Ensure that each team member can successfully log in and view their own user dashboard
Expected Results:	Each team member should successfully log in and can see their own user dashboard
Result:	-
Reason for Failure:	-
Remarks:	Successfully completed

9.2.1 Testing Remarks

Remark Id	Description

10. References

<https://doi.org/10.1016/j.rser.2015.12.001>

<https://doi.org/10.1109/isgt.2015.7131868>

<https://doi.org/10.1109/tel-net.2017.8343593>

<https://pages.cs.wisc.edu/~akella/CS744/F17/838-CloudPapers/Kafka.pdf>

<https://www.eia.gov/opendata/browser/electricity/electric-power-operational-data?frequency=monthly&data=ash-content;consumption-for-eg;consumption-for-eg-btu;consumption-uto;consumption-uto-btu;cost;cost-per-btu;generation;heat-content;receipts;receipts-btu;stocks;sulfur-content;total-consumption;total-consumption-btu;&start=2020-01&end=2022-12&sortColumn=period;&sortDirection=desc;>

Analysis Of Energy Consumption Patterns In The United States	Version: 1.1
Project Plan	Date: 05/05/2023