

Ethical Data Science Capstone
DS496 Capstone Project Final Report
Quentin McIntosh-Giles, Hannah Hodges,
Jose Ojea, and Maria Andreina Sisco

FutureMakers, Variable Management and Reporting

Abstract: This project involved the development of a scalable dashboard designed to monitor educator engagement with FutureMakers' Sparks kits. In the absence of real-world data, a relational database was constructed using mock data generated from entity relationship diagrams. A dynamic reporting tool was then implemented using Streamlit to enable interactive exploration of key engagement metrics. The resulting prototype is fully operational and prepared to support real-time analysis as live data becomes available.

Acknowledgments

We would like to express our sincere gratitude to Matt Barinholtz, CEO and founder of FutureMakers, for his support, guidance, and the opportunity to collaborate on this project. We also extend our thanks to Liam MacKinnon for his assistance with backend development and overall support. Finally, we are grateful to Dr. Kevin Drummey and Professor Jake London for their valuable input throughout the project.

Project Description

FutureMakers is an education-focused organization that equips educators and learners with hands-on resources, such as its *Sparks* kits, to foster creativity, engineering, and design thinking. As the organization continues to grow, the need for data-driven insight into client engagement has become increasingly important. To support this need, our project aimed to develop a minimum viable product (MVP) that enables FutureMakers to track and analyze educator interaction with their Sparks kits. Matt approached us with a clear objective: to better understand how clients use their products by analyzing key metrics such as resource engagement, training activity, and platform usage. In addition to building a reporting tool, we were also tasked with designing the infrastructure required to collect and store this data for future analysis.

To guide our work, we developed several core research questions:

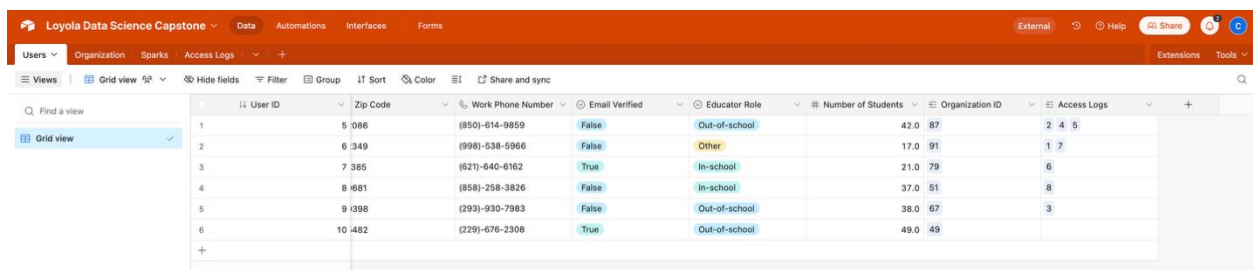
1. How can we design a reliable system to track educator engagement across platforms?
2. What templates and frameworks best support standardized event tracking for educator activities?
3. How can reporting dashboards be built to support real-time analysis at MVP launch?

With these guiding questions, we partnered with Matt and the FutureMakers team to design and implement a functional prototype capable of meeting the organization’s immediate needs and scaling for future use.

Data and Challenges

One of the key challenges our group encountered early in the project was the absence of a centralized and structured dataset ready for analysis. Initially, we anticipated that FutureMakers would have an established database that we could work with to uncover insights and trends related to user interactions. However, our understanding evolved after our first meetings, when we learned that the current data infrastructure was still in its early stages and not yet configured to support comprehensive analysis.

As a result, our project scope shifted from immediate data analysis to first building the necessary foundation for long-term data collection and organization. This began with learning Airtable, the platform FutureMakers uses for managing its information. Although we had limited prior experience with Airtable, we took time to explore its features, limitations, and best practices. We examined how Matt had structured his existing datasets, which included information on users, organizations, Sparks kits (the core educational resources), and some access logs. While these tables provided a valuable starting point, we identified opportunities to enhance consistency, completeness, and relational integrity to better support analytical needs moving forward.



The screenshot shows an Airtable interface for a table named "Users". The table has the following columns: ID, User ID, Zip Code, Work Phone Number, Email Verified, Educator Role, Number of Students, Organization ID, and Access Logs. The data is displayed in a grid view with 6 rows of user information.

	ID	User ID	Zip Code	Work Phone Number	Email Verified	Educator Role	Number of Students	Organization ID	Access Logs
1		5 086		(850)-614-9859	False	Out-of-school	42.0	87	2 4 5
2		6 349		(998)-538-5986	False	Other	17.0	91	1 7
3		7 385		(821)-640-6162	True	In-school	21.0	79	6
4		8 481		(858)-258-3826	False	In-school	37.0	51	8
5		9 1398		(293)-930-7983	False	Out-of-school	38.0	67	3
6		10 482		(229)-676-2308	True	Out-of-school	49.0	49	

The left screenshot displays the 'Sparks' view in the Loyola Data Science Capstone application. It features a table with columns: Spark ID, Name, Grade Level, and Price per Learner. The table lists 18 items, including 'Art Machines', 'Blenders', 'Crankle Contraptions', 'Draebots', 'Electric Vehicles', 'Flying Machines', 'Hypnotizers', 'Linkages', 'Mazes', 'Monster Claws', 'Monster Mouths', 'Scene Machines', 'Straw Structures', 'Strawbots', 'Scene Machine', 'Art Machines 2', 'Electric Cars', and 'Mini Crankles'. Each item is associated with a specific grade level (e.g., Pre-Ki, Grade 2+, Grade 4+, Grade Ki) and a price of 5.95.

The right screenshot displays the 'Access Logs' view in the same application. It features a table with columns: Access ID, User ID, Spark ID, Timestamp, and Viewed Slideshow. The table shows a list of user interactions, including the Access ID, User ID, Spark ID, Timestamp, and whether the slideshow was viewed (indicated by a checkmark).

We spent a significant portion of our project time restructuring these data buckets in a way that made them more relational and analytical. Our group also worked closely with Matt to understand the types of questions he wanted to answer about user engagement. He was interested in knowing how educators interact with Sparks, which parts of the site receive the most engagement, and how those interactions relate to educator understanding and student outcomes. In order to answer these types of questions, we needed to think critically about how data should be captured, and which types of user actions needed to be tracked. This led us to analyze the Future Makers website to identify potential listening points, or moments where data about user behavior could be collected. These included actions like viewing a slideshow, watching a tutorial video, downloading lesson materials, or booking a support session.

To prepare the variables used in our analysis, we began by constructing an initial entity relationship diagram (ERD) to visualize the ideal data structure for Future Makers, and to facilitate understanding for Matt. This first draft (see Appendix, "Entities Diagram - First Draft") included entities such as USER, ORGANIZATION, PRODUCT, ORDER, and SATISFACTION. The intention was to treat educational resources as products, and track usage through orders and satisfaction feedback. However, after several iterations and discussions, we realized that this schema did not align well with how educators actually interact with the Future

Makers platform. As a result, we made substantial adjustments. The PRODUCT entity was renamed to Spark to reflect the actual terminology used by the client, and the ORDER relationship was eliminated entirely, as it did not represent a meaningful or realistic interaction. Similarly, the SATISFACTION entity was restructured into what is now ACCESS LOG, a more accurate and flexible representation of user engagement data.

The final version of our ERD (see Appendix, "Entities Diagram - Final Draft") reflects these changes and served as the foundation for our mock database. The updated schema consists of four primary entities: USER, ORGANIZATION, SPARK, and ACCESS LOG. Each entity includes a clearly defined set of attributes, such as educator role, Spark grade level, session length, and interaction type, that we determined based on real use cases and Matt's reporting needs. Relationships such as "works at" and "contains" were established to define how users are connected to organizations, and how their interactions with Sparks are logged. The Access Log table became the central hub for capturing key actions like viewing slideshows, booking support sessions, and using the AI Playbook Maker, all measures that the current website was already gathering or could be easily implemented. This relational model allowed us to create a normalized, scalable dataset with realistic usage patterns. From there, we wrote Python scripts to populate each table with mock data aligned to these entities and their connections, giving us a comprehensive simulated environment for testing analytical queries and visualizations.

The absence of actual user interaction data was one of the most difficult hurdles we faced. Because the infrastructure for data collection had not yet been implemented on the live website, there were no records of how educators were currently using the site. This created a major obstacle for testing and validating any analytical solutions we hoped to build. To overcome this, we generated thousands of rows of mock data that simulated realistic user interactions. We created fake users, organizations, Sparks, and detailed access logs to mimic

potential site behavior. This allowed us to test the database structure, experiment with queries, and prototype the analytical tools that could be used once real data becomes available.

While the lack of real data was a challenge, it ultimately gave us the opportunity to contribute something much more foundational. Instead of simply delivering analysis, we designed and implemented a system that will allow Future Makers to capture, store, and analyze meaningful user engagement data moving forward. This system lays the groundwork for long-term insights and reporting. Our mock data served as a placeholder, but the structure we built is ready to accommodate actual usage data as the platform continues to grow. We are confident that the infrastructure we established will enable Future Makers to track educator engagement, identify trends in resource use, and support data-informed improvements to the Sparks and overall platform experience.

Analytical Methods

The technical development of the FutureMakers dashboard began with implementing a data-driven web application using Streamlit. Our primary goal was to build a tool that could visualize and analyze educator engagement with the Sparks kits. To do this, we created a system that supports interactive data uploads and dynamic visualizations based on user-provided CSV files.

The code snippet initializes this system by importing essential Python libraries—such as `'pandas'` for data manipulation, `'plotly'` and `'seaborn'` for rich data visualizations, and `'streamlit'` for building the web interface. These tools allowed us to create a clean and responsive application that educators and administrators can use without needing technical expertise.

```
# Libraries
import streamlit as st
import pandas as pd
import plotly.express as px
import seaborn as sns
import matplotlib.pyplot as plt
from datetime import datetime
```

We configured the dashboard layout with `st.set_page_config`, giving it a professional and readable appearance. Next, we built a sidebar interface for users to upload data files relevant to engagement tracking: access logs, user data, organizational context, and Sparks resources. This setup allows the application to process real-world usage data uploaded directly by clients, making the dashboard highly adaptable.

```
# Function for generating the Account Report in Streamlit
def from_code_account_report(access_logs, users, organizations, sparks):
```

The foundational function `from_code_account_report` was designed to tie these elements together. It processes uploaded data and prepares it for reporting—paving the way for real-time insights into how educators interact with FutureMakers' offerings. By handling these operations dynamically, the dashboard empowers Matt and his team to evaluate platform adoption, identify active users, and gain a better understanding of their impact—key elements in shaping the future direction of the Sparks program.

```
# --- Streamlit App Setup ---

# Set page configuration
st.set_page_config(page_title="Spark Engagement Reports", layout="wide")

# Sidebar inputs for CSV uploads
st.sidebar.header("Upload CSV Files")
access_logs_file = st.sidebar.file_uploader("Upload access_logs.csv", type=["csv"])
users_file = st.sidebar.file_uploader("Upload users.csv", type=["csv"])
organizations_file = st.sidebar.file_uploader("Upload organizations.csv", type=["csv"])
sparks_file = st.sidebar.file_uploader("Upload sparks.csv", type=["csv"])
```

In addition to developing a functional dashboard, one of the most substantial parts of our contribution involved applying core data science concepts to generate visual insights from the mock data we built. We produced five total reports—three customer-facing and two internal. Each report generates downloadable tables and graphs based on selected parameters like organization, user, and date range.

Spark Engagement Summary

	Spark Name	User Sessions	Percent Resources Used	Timestamp
0	Scene Machine	1	14.2857	2025-04-01
1	Scene Machine	1	14.2857	2025-04-02
2	Art Machines	1	14.2857	2025-04-04
3	Mazes	1	14.2857	2025-04-04
4	Art Machines	1	14.2857	2025-04-05
5	Drawbots	1	28.5714	2025-04-05
6	Mazes	1	42.8571	2025-04-05
7	Drawbots	1	14.2857	2025-04-06
8	Blendies	1	42.8571	2025-04-10
9	Crankie Contrapti	1	28.5714	2025-04-10

Select an Account (Organization)

Torres Ltd

Available Date Range: 2025-04-01 to 2025-04-21

Start Date

2025/04/01

End Date

2025/04/21

This modular structure ensures the dashboard can be used flexibly across different schools, sites, and educator types, supporting both granular and summary-level analysis. The backend logic powering these reports was written in Python, leveraging pandas for data aggregation and plotly for interactivity, as stated before.

Three graphs in particular highlight the power of our approach. First, the User Activity Timeline (see Appendix: Figure 3) is generated in the Individual User Report. It provides a visual map of an educator's journey across the platform, showing the types of interactions (e.g., watching a video, booking a session) that occurred over time. The timeline format is useful for identifying patterns in how educators engage with Sparks, as well as potential drop-off points in usage. Second, the Session Length Distribution per Spark (Figure 5, from the Account Report) illustrates variation in session duration across different Sparks. While no conclusions can be drawn from the mock data, the visualization showcases how real-world usage patterns could be tracked to determine which kits are more engaging or require more time investment. Lastly, the Accesses Over Time graph (Figure 6, from the Site Report) tracks aggregate usage volume within a selected date range, allowing administrators to observe platform activity over days or weeks. Spikes in this line graph could correspond with sessions or Spark releases in a live environment.

More graphs were generated for each report, and the demo of how it all looks like live can be found in Appendix 4, Demo. All of these graphs are powered by filtering the uploaded access logs, which are parsed in real-time using functions within the dashboard. This allows the same underlying system to drive multiple reports with different user perspectives. Importantly, although these visuals are built using mock data, they fully demonstrate how real metrics will be presented and used in future decision-making. For example, the Resource Engagement Summary table (included in the report body before) lets users track resource consumption per Spark over time. Similarly, dropdown selections allow users to isolate an individual educator, a specific school, or a date range of interest. These features were designed to accommodate FutureMakers' evolving data needs, including potential integrations with Airtable or other data storage services. Overall, we applied key data science principles, including ETL (extract, transform, load), data

cleaning, relational modeling, and visualization, to create a functional prototype that can scale as real usage data becomes available.

Conclusions

Through our analysis and dashboard development, we were able to simulate and test the reporting system that FutureMakers can use moving forward. Although we worked with mock data, the structure of our analysis allowed us to validate that the backend logic and visualizations performed as intended. For each of the five reports—Account, Site, Individual User, Sparks, and Resource Type—we confirmed that data could be dynamically filtered by date, organization, and user to generate relevant tables and graphs. These outputs responded accurately to dropdown selections and reliably reflected the conditions set by the user, proving that our ETL and filtering pipeline was functioning correctly.

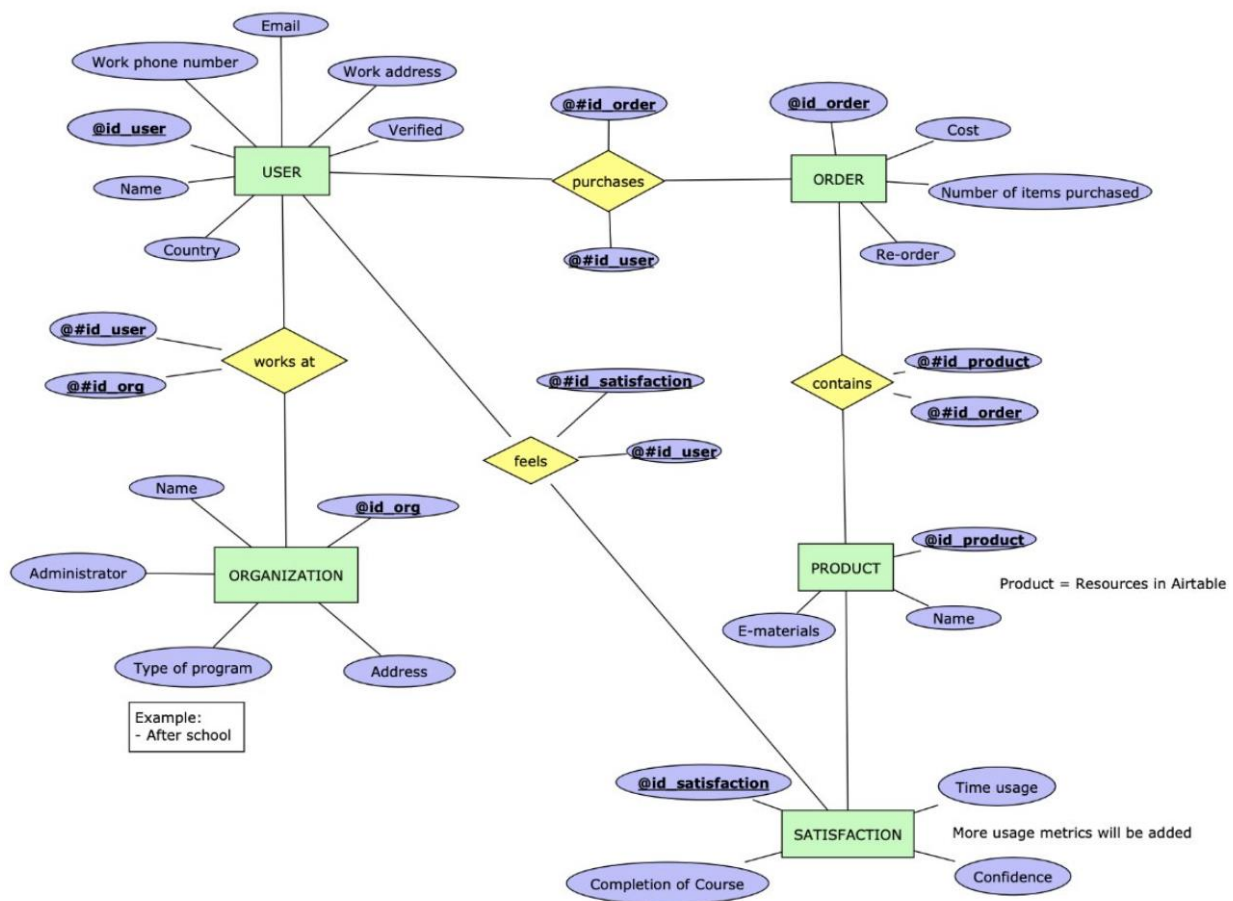
More specifically, we verified that our system could compute critical engagement metrics, such as session counts per Spark, average session lengths, and the percentage of resource types accessed. We confirmed that different user roles (e.g., educators, administrators) could be associated with their respective sites and that those relationships were correctly represented in the final outputs. We also demonstrated that data from multiple entities (users, organizations, access logs) could be joined, transformed, and visualized seamlessly within a single report. While we were not able to extract insights from actual usage behavior, our results show that the system is ready to do so as soon as real data becomes available.

The scope of our project was not to bring to light information from the data but to formulate an interface for the client to create dashboards as easily as possible. With the help of python and the Streamlit dataframe we can utilize the data created from the features in Airtable to pull specific columns from. Although the data itself did not reveal any insight as it was mock data, this project did reveal different metrics that we currently did not think about tracking for the client. There were several aggregated feature that can be created through Airtable which will be particularly useful when the real data starts coming in.

Appendix

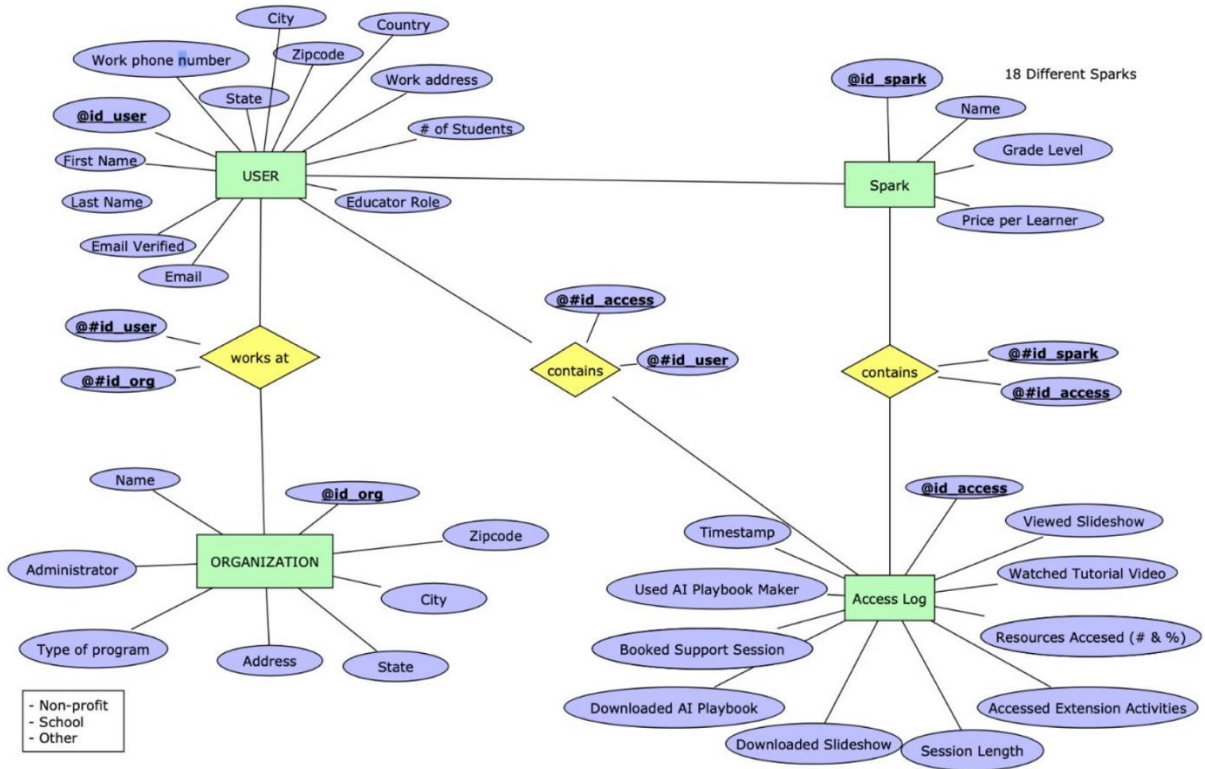
1. Entities Diagram, First Draft:

ENTITIES DIAGRAM FOR FUTUREMAKERS DATABASE



2. Entities Diagram, Final Draft:

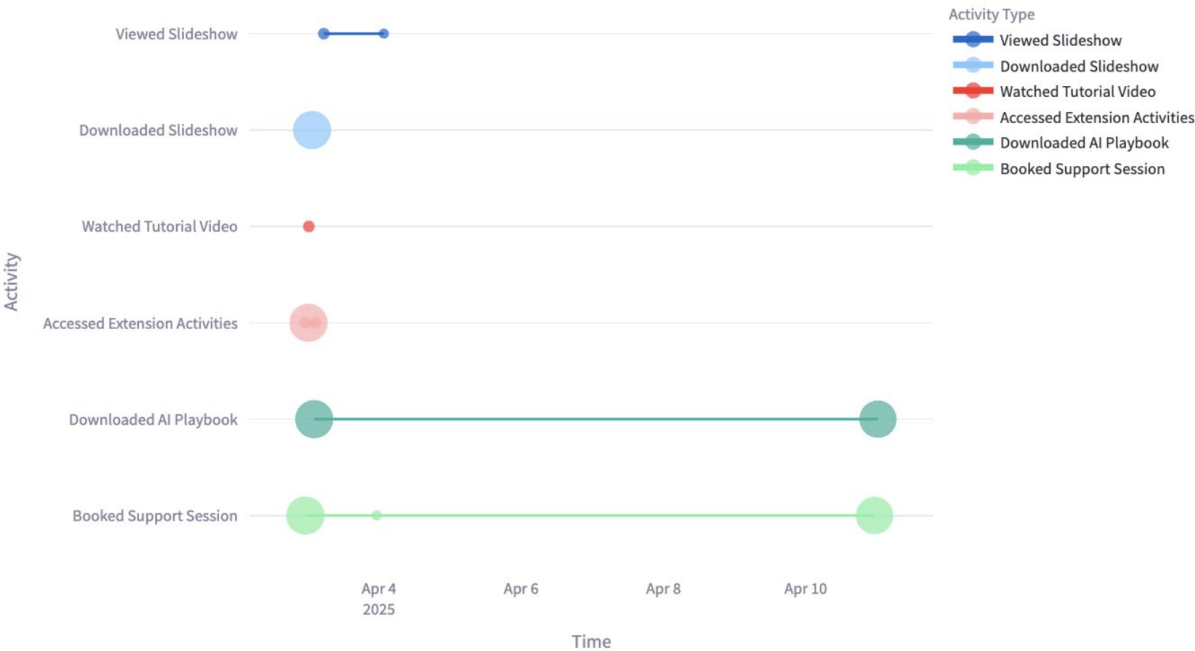
ENTITIES DIAGRAM FOR FUTUREMAKERS DATABASE



3. User Activity Timeline

User Activity Timeline

Journey of Christopher Barnes

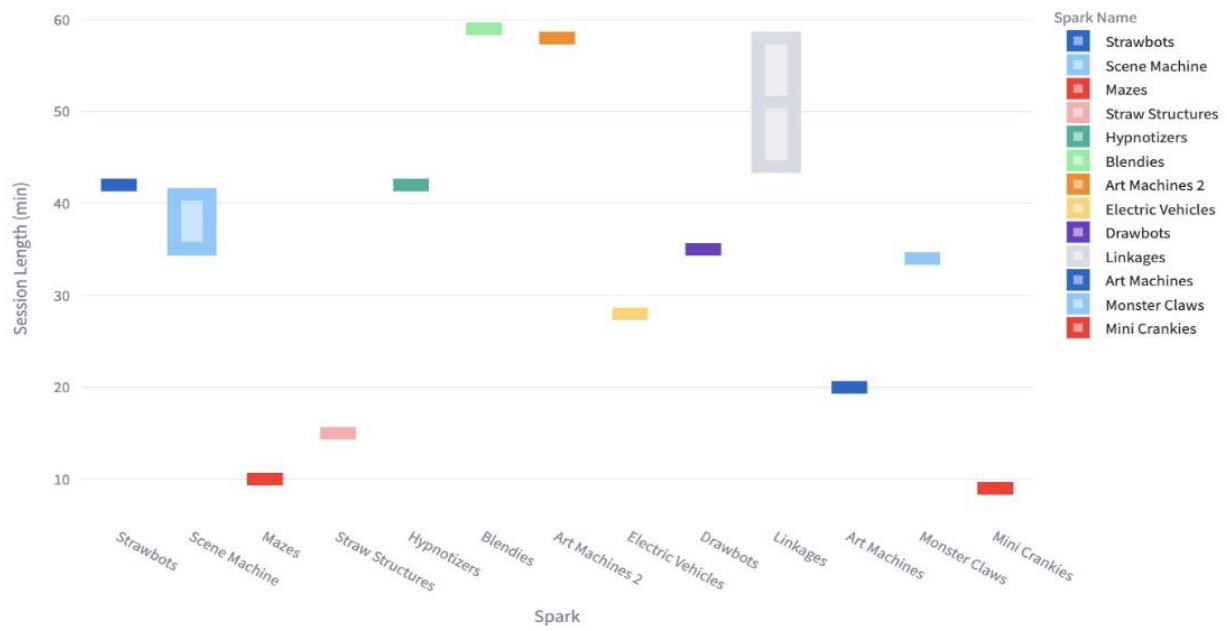


4. Demo:

<https://youtu.be/Av8QAjIctN0>

5. Session Length Distribution per Spark

Session Length Distribution per Spark



6. Accesses Over Time graph

