

Phase 1 Project Launch

May 25, 2023

// **FLATIRON SCHOOL**



Agenda

1. **Building a Professional Data Science Portfolio**
2. **Project Repo and Deliverables**
3. **Why Group Projects?**
4. **Group Project Best Practices**
5. **Group Assignments and Schedule**

Building a Professional Data Science Portfolio



Medium



GitHub

Project Repo and Deliverables



Important Canvas Links

- [Project Assignment](#)
 - Includes rubric at the bottom.
 - Project is Pass/Not Pass
 - Only 1 submission per group
- [Checklist](#)
 - Provides added detail to the rubric

Deliverables



**Github
Repository**

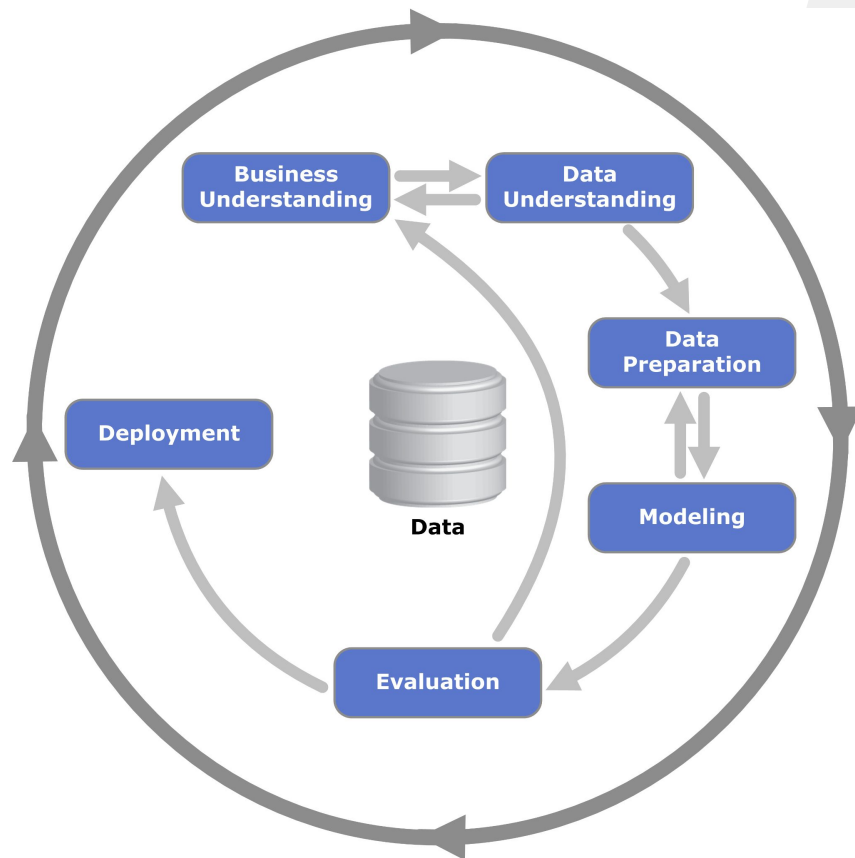
**Technical
Notebook**

**Non-Technical
Presentation**

**Interactive
Dashboard**

CRISP-DM

While creating each deliverable, think about the **CRISP-DM** process

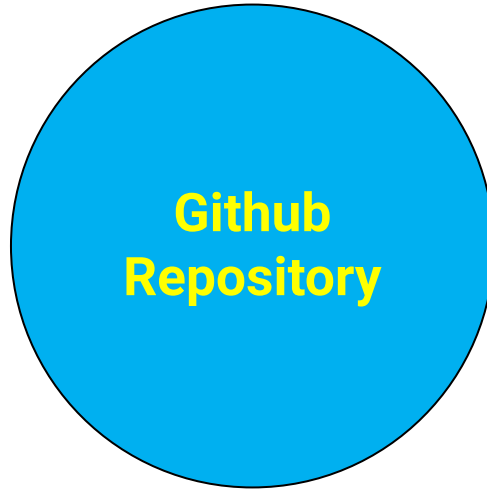


CRISP-DM

- You can think of deployment as your completed deliverables and three recommendations.



Deliverable 1



Deliverable 1: Github Repo

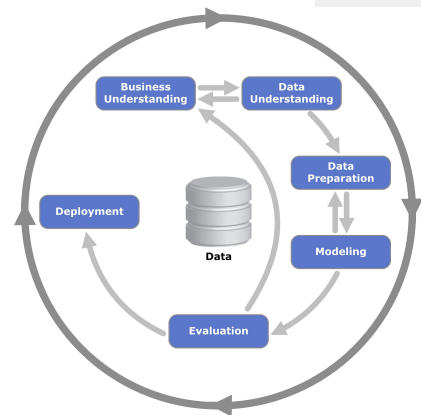
- Where your project lives and grows.
- You will provide links to these repos to employers to demonstrate your coding, analytical, and problem problem solving skills.
- Fork a template repository or start from scratch.
 - [Sample Repository and template](#)

Github Repo Must Haves

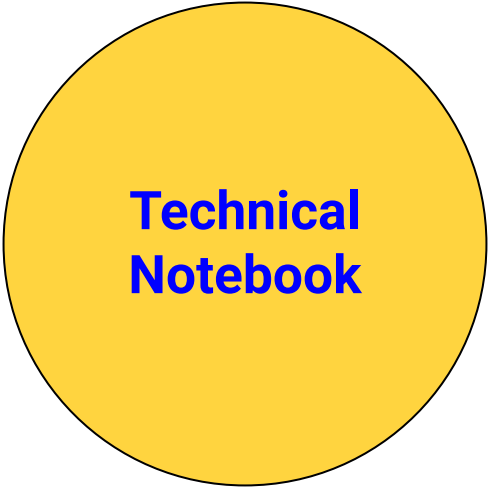
01	README.md	<ul style="list-style-type: none">• See next slide
02	Commit History	<ul style="list-style-type: none">• Commit history with clear messages• Contributions from all team members
03	Organization	<ul style="list-style-type: none">• Clear folder structure• Clear names of files and folders• Easily located technical notebook and presentation linked in the README.md
04	Notebooks	<ul style="list-style-type: none">• Final technical notebook that runs without errors• Working notebooks from individual team members
05	.gitignore	<ul style="list-style-type: none">• Ignores large files and pesky things like .ipynb_checkpoints• Github python gitignore

Github Repo: README.md

- **General Overview**
- **Business Understanding**
- **Data Understanding**
 - Include stakeholder and key business questions
- **Data Analysis**
 - Describe *interesting* techniques used to prepare/filter data
 - Written interpretation of statistics which inform business questions
 - **An Interactive Dashboard** with appropriate visuals
- **Conclusion**
 - Summary of conclusions including 3 relevant findings



Deliverable 2

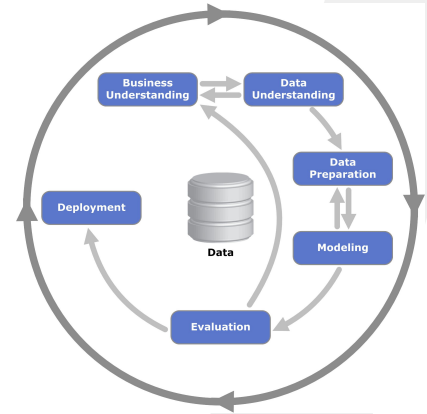


Deliverable 2: Technical Notebook

- **Brings together** the important components of each group member's individual notebooks.
- Blends code, comments, markdown, and visualization to tell the **full story of your project**.
- Notebook should be free of errors and runs from top to bottom.
- [Sample Final Notebook](#)

Deliverable 2: Technical Notebook

- Use **CRISP-DM** to help organize your final notebook
- Use **markdown headers** to divide your notebook into sections
- O.K. for content to overlap with the README.md and non-technical deck.



Note on Jupyter and Github

- Your final notebook deliverable is one .ipynb file with the important contributions of your group members.
- However! Create **individual notebooks** to work on as you go to avoid merge conflicts.

Deliverable 3

**Non-Technical
Presentation**

Deliverable 3: Non-Technical Slide Deck

- Used during 5 minute presentation
- Intended for non-technical audience
- Professional style
 - Light on text
 - Legible and labeled visualizations
- Example Slide Template END of this DECK

“HOW TO DO A DS PRESENTATION” – Jelly Style

- Link: Slidedeck for [“How to do a presentation”](#)
- REVIEW THIS DECK IN ITS ENTIRETY

Deliverable 3: Non-Technical Slide Deck

Tell a story:

Beginning

- Overview
- Business Understanding
- Stakeholder
- Key Business Questions

Middle

- Data understanding and analysis
- Key statistics supporting findings
- Key visualizations supporting findings

End

- Recommendations
- Next steps
- Thank you slides

Why Group Projects?





#SchittsCreek



I'M FEELING THIS, LIKE, DEEP
ACHING SENSE OF DREAD

For some of you, this might be your immediate reaction to the phrase “group work”...



WHEN **ONE** OF
US SHINES

#SCHITTSCEEK

Pop

...but hopefully we can
convince you to feel more like
this

Group Project Best Practices

Remote Edition

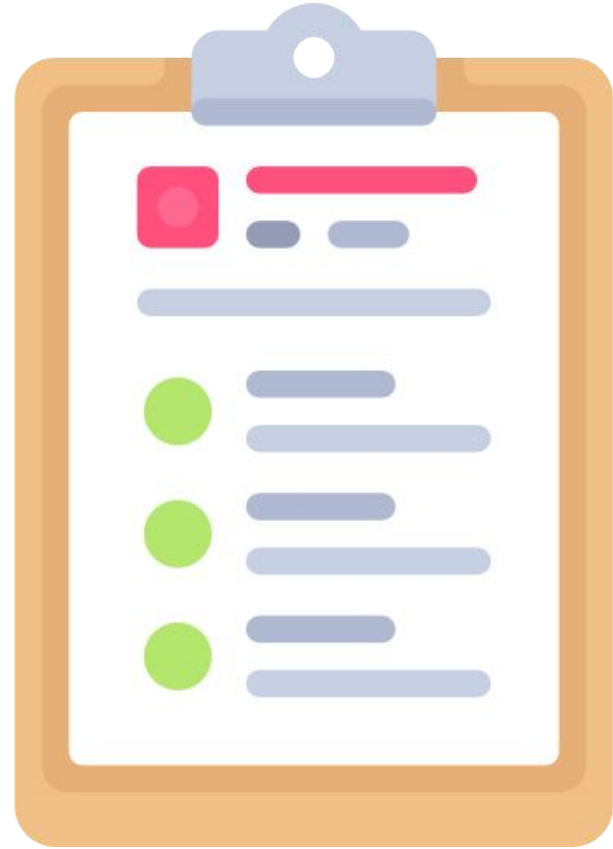


Group Project Best Practices

1. Get to Know Your Group Members
2. Define Individual Project Contributions
3. Meet Regularly
4. Communicate Actively, Clearly, and Transparently

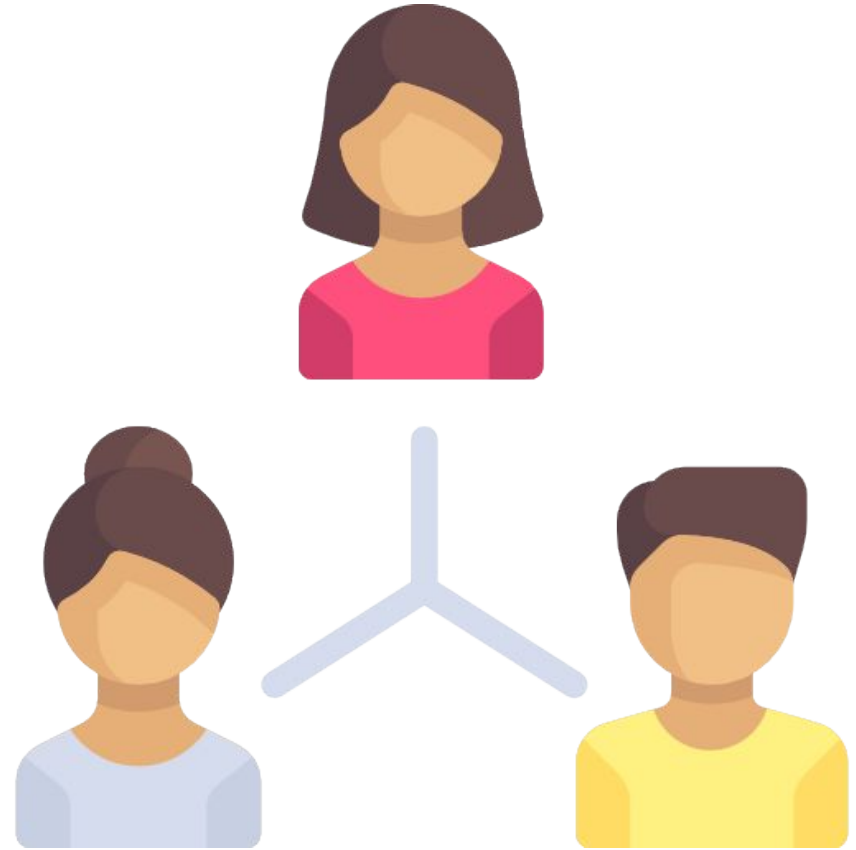
Credit to Dr. Grace Oh, University of Illinois College of Education ([website](#))

Icons made by [Freepik](#) from [Flaticon](#)



Get to Know Your Group Members

1. Expectations
2. Experience
3. Schedule



Define Individual Project Contributions

1. For every project, make sure there is some specific piece that you can individually claim responsibility for
2. Try to divvy up contributions as early in the process as possible
3. The final deliverables should combine the individual contributions
 - E.g. the final notebook should contain contributions from everyone



Group Assignments and Schedule



Project Groups.

- Group 1: Morgan (GL), Simon (PL/Tab), Justin (TL)
 - Group 2: Maddie (GL), Raj (PL/Tab), Lillian (TL)
 - Group 3: Jordan (GL), Jimmy (PL/Tab), Stuart (TL)
 - Group 4: George (TL/GL/PL)
 - Group 5: Yun (TL/Tab), Steven (PL), Sean (GL)
-
- GL = GitHub Lead, PL = Presentation Lead, TL = Tech Lead, Tab - Tableau Lead

Roles - Tech Lead

- Tech Lead

- Gains input from team on what Master Dataset should be created and spearheads the creation of that Master Dataset (if applicable)
- Provides technical support (aka troubleshooting code) for teammates experiencing coding issues
- Gains input from team on what machine learning models should be utilized in solving client problems and decides the modeling workflow
- Ensures that there is no data leakage!
- Ensure that the code is readable, commented, concise, and reproducible
- Consolidates all code into a final jupyter notebook
- **DOES NOT code everything but decides who is doing EDA, Model Building, etc. (everyone should have a coding part; however, it is up to the Tech lead to vet all code)**

Roles - GitHub Lead

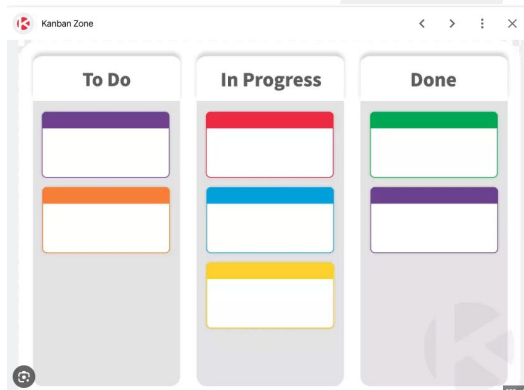
- GitHub Lead
 - Creates the project repo
 - Adds dataset to the project repo
 - Creates branches for individual groupmates (if applicable)
 - Creates README
 - Creates Repository Navigation within the README
 - Ensures folder names and file names are descriptive
 - **Ensure commits are descriptive**
 - **Creating a .gitignore**
 - Helps troubleshoot GitHub issues groupmates may have

Roles - Presentation Lead

- Presentation Lead/Tableau Lead
 - Creates the presentation template and layout
 - Consolidates technical findings into client-facing slides
 - Works with team to decide presentation order and flow
 - Ensures slides have minimal text and appropriate visuals
 - **Ensures technical jargon and metrics are LEFT OUT of the presentation**
 - Ensures the group practices the presentation and gives actionable feedback to team on how to improve presentation skills
 - Spearheads answering audience questions
 - Ensures presentation is approximately 5 mins
 - Interactive dashboard is sound

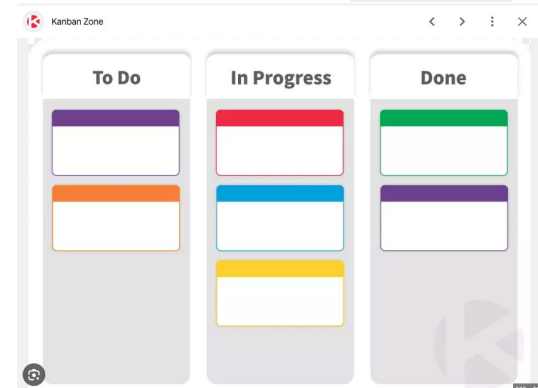
Steps to Get Started

- **READ** Project Description: [Project Description](#)
- Create a GitHub Repository (add each other as collaborators and create branches)
- Create a notebook on EACH branch so everyone can explore the data
- Set times for team meetings and think about the **project plan**
- Create a [Notion](#), [Trello](#), or [Monday](#) board to assign responsibilities/tasks (Kanan Board)



REMEMBER:

- **EVERYONE** in the group **MUST** use the same data for the analysis (aka if you filter the data everyone needs to use the same filtered data)
- SET EXPECTATIONS and stick to them
- IT IS A GROUP project so the analysis should **FLOW**
- TELL A STORY and use data to back it up



Project Events

- **Friday:** Project work
- **Wednesday:** Group Check-ins
- **Thursday AM:** Practice Presentations
- **Friday AM:** Final Presentations
- **Friday 12 midnight EST:** Submit link to repo and completed deliverables

Example Phase 1 Projects (OLD but Useful)

- <https://github.com/Baekho4245/microsoft-movie-analysis>
- <https://github.com/JulianWard147/New-Film-Studio-at-Microsoft>
- https://github.com/ewjohn127/dsc_project_1

Questions?

Appendix

Skeletal Slide Deck Template



Microsoft Revenue Analysis

Jan 27, 2023



TEAM



Name: Kirby Dream
Role: Technical & Version Control Lead
Contact: [githubkirby](#)



Name: Kirby Popstars
Role: Business Liaison & Presentation Lead
Contact: [githubkirbypop](#)

Agenda

**Business
Problem**



**Data
Overview**



Analysis

Recommendations



**Future
Steps**

Bottom Line (or Main Finding)



COMEDY

**10x increased
revenue**



The image features a white background with two green geometric shapes in the corners. The top-left corner has a green shape that looks like a stylized 'L' or a corner piece. The bottom-right corner has a similar green shape, also a stylized 'L' or corner piece.

BUSINESS PROBLEM

Business Problem

- Microsoft breakthrough movie industry
- Highest Revenue Genre
- Highest Rated Actor/Actress
- Best Movie Studio partnership

Image

Image

Image

Business Problem

Data Overview

Analysis

Reccs.

Future Insights

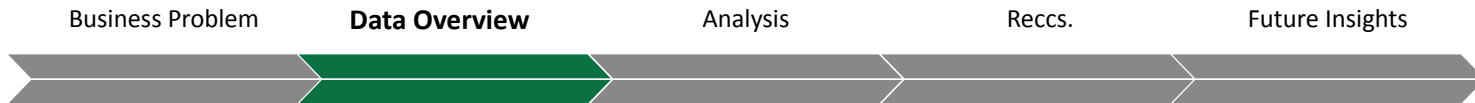




DATA OVERVIEW

Data Overview

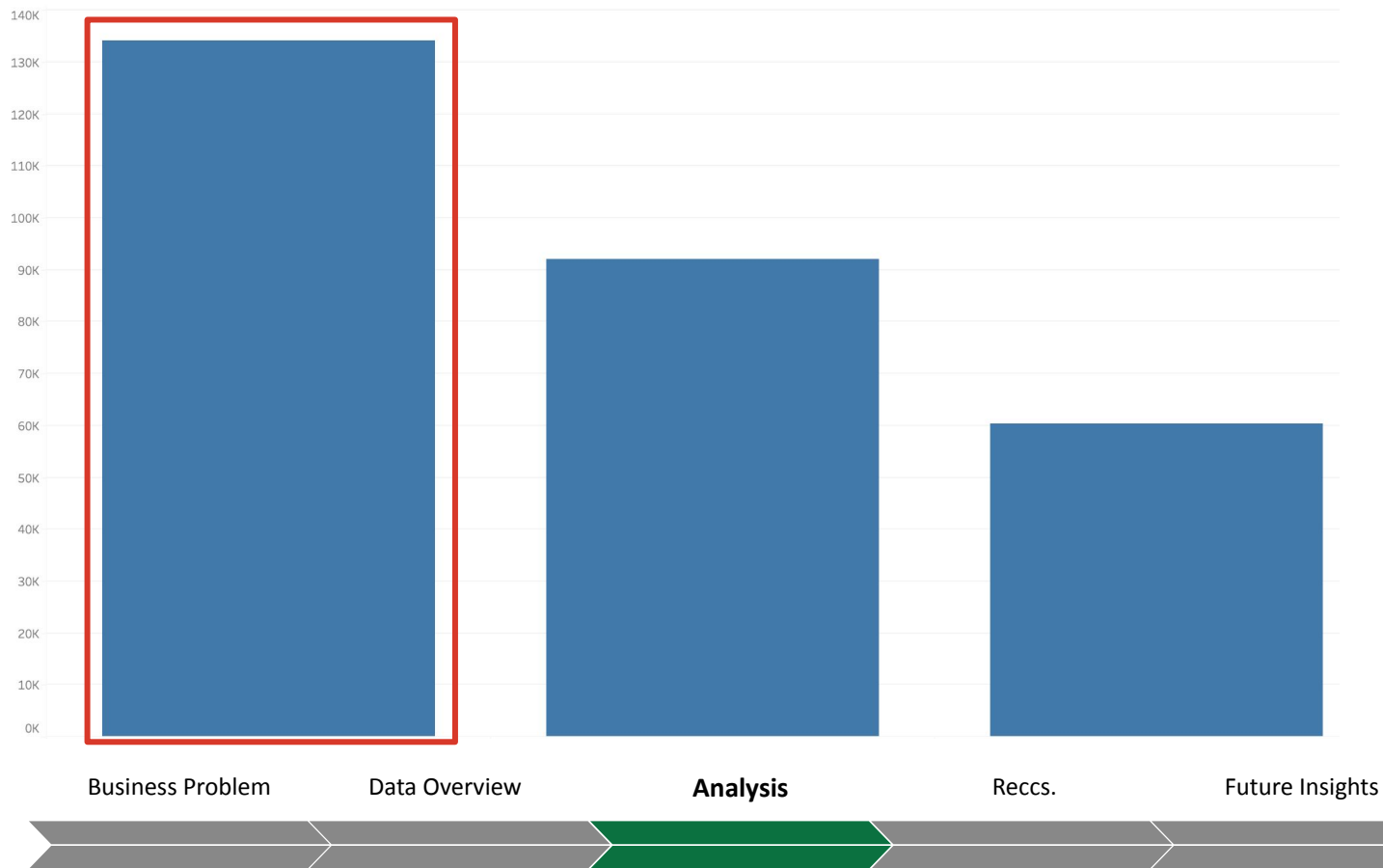
- Where did the data come from?
- Time range of the data?
- How many data points? (40,000 movies)
- Does it span across multiple locations (natl vs. international, South America etc.)
- LIMITATIONS (values not adjusted for inflation?, no information of budget?)



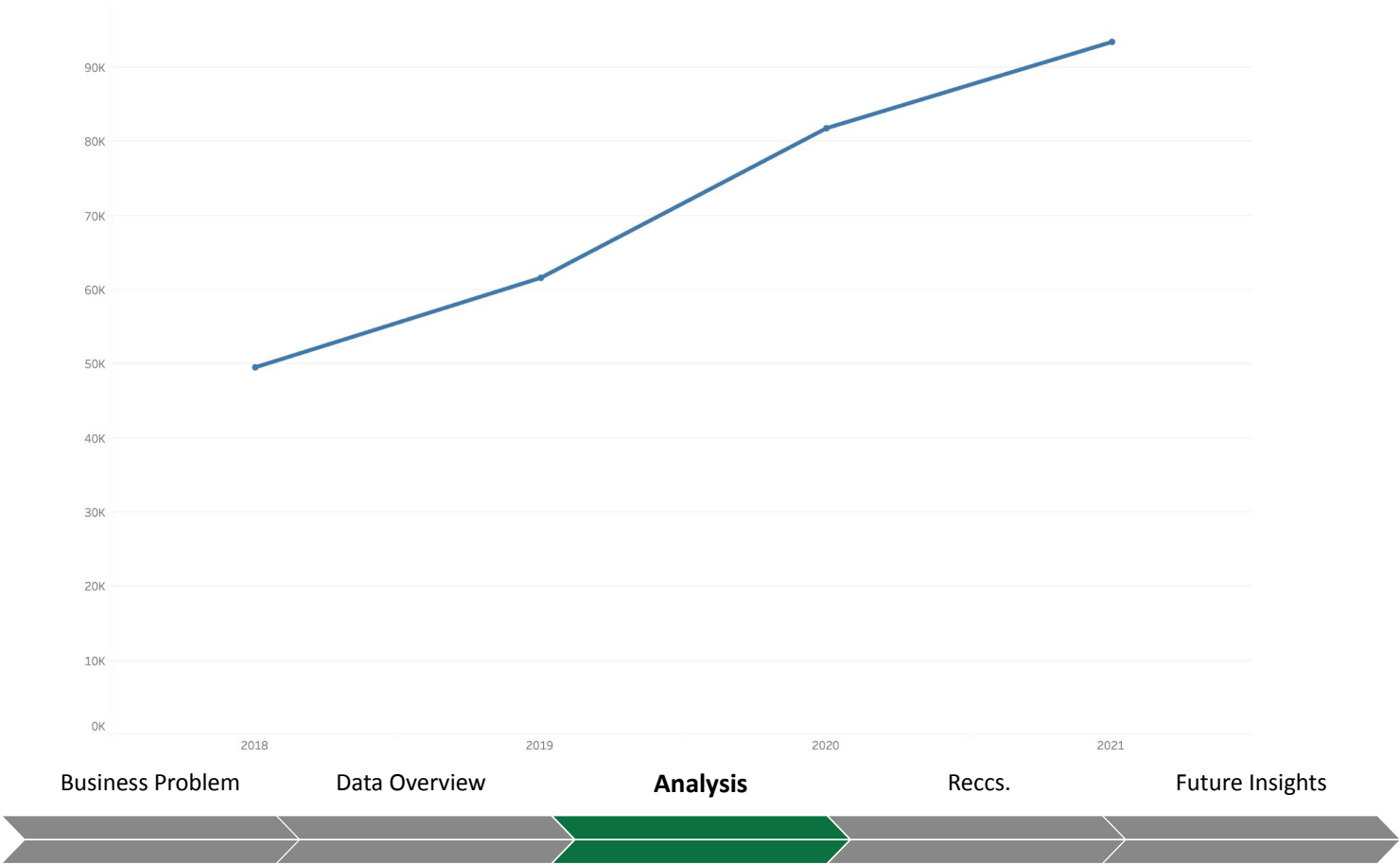
The image features a white background with green geometric shapes in the corners. In the top-left corner, there is a green shape that looks like a stylized 'L' or a corner piece. In the bottom-right corner, there is a similar green shape, also a stylized 'L' or corner piece.

ANALYSIS

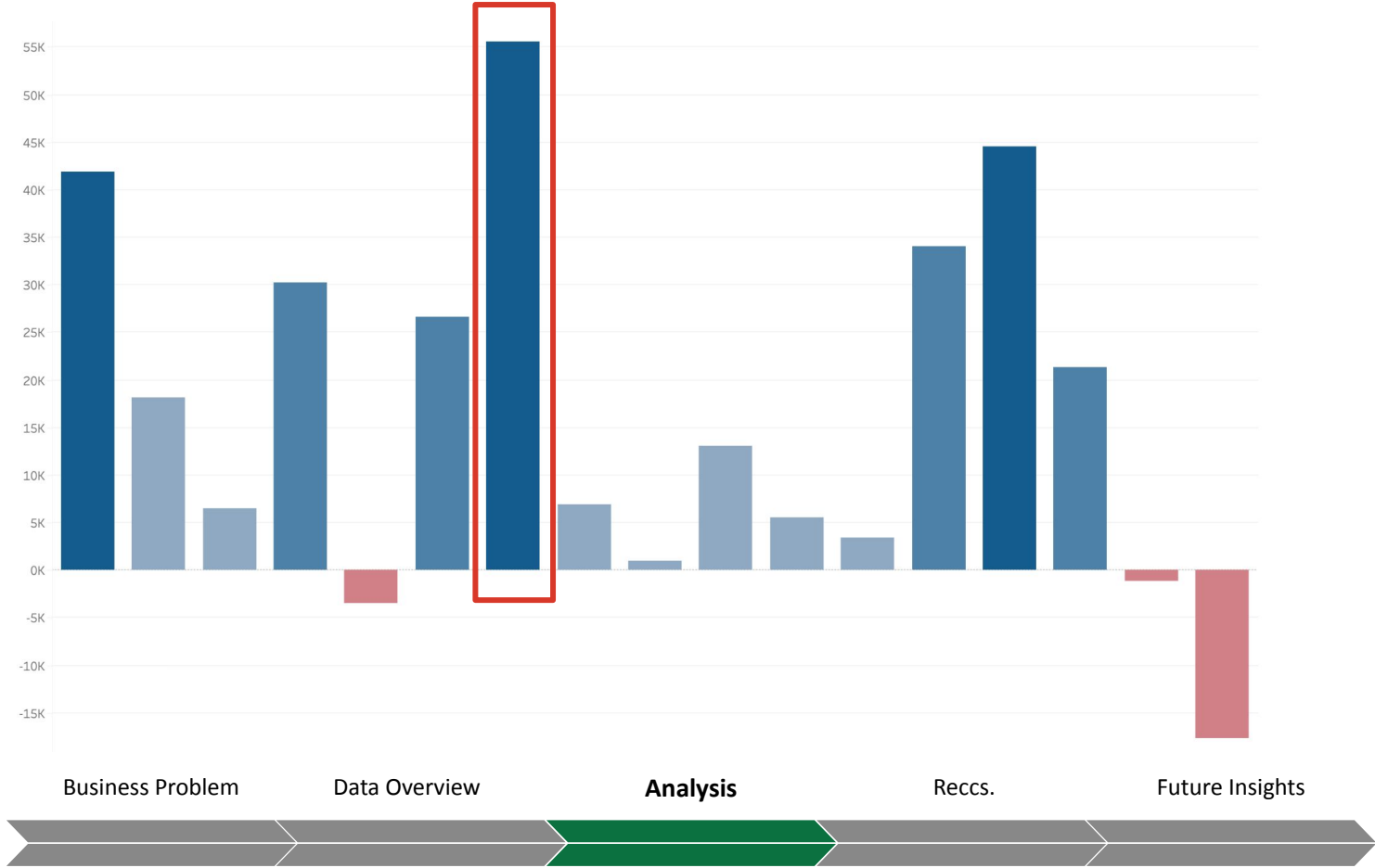
Top Genre with Highest Revenue is Comedy



Upward Trend in Revenue Across Time for Comedies



Top Actress Featured in Highest Revenue Comedy is Jelly Spratley



Dreamworks has Delivered the Most Recent Highest Revenue Comedy





RECOMMENDATIONS

Dreamworks has Delivered the Most Recent Highest Revenue Comedy

Go for Comedies

Jelly Spratley as Actress

Learn from Dreamworks



Business Problem

Data Overview

Analysis

Reccs.

Future Insights



The image features a white background with two green geometric shapes. One shape is in the top-left corner, and the other is in the bottom-right corner. Both shapes are composed of several rectangular blocks of varying sizes, creating a stepped, architectural appearance.

Future Insights

When Should Microsoft Launch Their Comedy...

- Analyze Seasonal Trends
- Studio Partnership is Feasible
- Marketing Strategies

Image

Image

Image

Business Problem

Data Overview

Analysis

Reccs.

Future Insights



TEAM



Name: Kirby Dream
Role: Technical & Version Control Lead
Contact: [githubkirby](#)



Name: Kirby Popstars
Role: Business Liaison & Presentation Lead
Contact: [githubkirbypop](#)



Questions?

Appendix