BREAK THROUGH TECH

AI Studio Project Scope and Deliverables

This document is designed to help your team understand, internalize, and align on the business context for your AI Studio project challenge and what is expected of you by your Challenge Advisor. Complete all 4 sections below as a team to the best of your ability. Then, review the document with your Challenge Advisor in your upcoming Sprint #1 meeting and make any additions or changes. Finally, submit your final Project Scope & Deliverables document by September 4th as part of your team's Sprint #1 deliverables.

Company / Organization	KPMG
Project Title	Charging into a Sustainable Future
Team Members	1. Alice Liu / CCNY / alice.liu37@gmail.com 2. Shoya Dixon / Queens College / shoya.dixon177@gmail.com 3. Rachel Tong / Cornell / rt387@cornell.edu 4. Melina Tsai / Cooper Union / tsaimelina@gmail.com 5. Wang Xiang / NYU / wx2008@nyu.edu 6. Ana Maria Rodriguez / Columbia / amr2343@columbia.edu
Challenge Advisor(s)	1. Jim Leach / Data Scientist, Manager / jimleach@kpmg.com
AI Studio TA	TBD (team can fill in later)
Final Presentation Date	TBD (team can fill in later)

PART 1: PROJECT OVERVIEW

Project Description

In your own words, what are you trying to accomplish? Give a concise goal statement.

Analyze data and create a model that provides a descriptive location on placing EV Charging Stations in the city of Dallas. The model is for the private sector.

Purpose of Project

Why is this project important? What impact will this project have on the business?

Increasing the amount of registered EV in Texas that will help the environment.

- Currently, only 1% of Texans' registered vehicles are EV.
- A factor that prevents customers from transitioning to electric, is the amount of EVCS that are available.

Expand the client base that KPMG can advise - private sector

- Give more specific advice for charging locations than is currently feasible

External Use Cases

Consider industry research and how your project might impact industry or the wider business ecosystem.

- This project would impact industries by:
    - EVCS for businesses can encourage EV drivers to spend more time and money in stores and restaurants, etc.
    - Businesses have incentives for installing charging stations. With the help of EV Connect, business owners stand to access a combined $2.6 billion in incentives through utility and local government programs.

- Attract more customers to purchase EV which will benefit EV manufacturing companies.
- Utilities providers will benefit from a drastic increase of electric usage by EV owners.

Ethical Considerations

What are some ethical concerns or considerations for your project and/or its impact?

We may access:

- EV users' demographic information
- Pairing of EV user and vehicle information by registration
  - potential increased surveillance
- EV users' charging location and time
- Household information in Dallas

Justice40: This project will make an impact in fulfilling Justice40 which states "40% of the overall benefits of certain Federal investments...must flow to disadvantaged communities." Investments include clean transit. The Bipartisan Infrastructure Bill targets Justice40 as the bill. Our goal with this project is to increase accessibility to EV charging stations (EVCS) in Texas which will come with the following benefits:

- Improving disadvantaged communities and encourage balanced growth in the region of Dallas

PART 2: PROJECT SCOPE

Project Requirements

At a high level, what is your Challenge Advisor expecting your team to deliver? Is there a specific type of model(s) that they would like you to test? (e.g. supervised, unsupervised, etc.) What is your team optimizing for? (e.g. accuracy, efficiency, no false negatives, etc.) What is the output of the agreed upon solution? Does your Challenge Advisor expect you to deploy a solution?

- There's no specific model to test. It's for us to decide what's the best one
- Since we're working with private sectors we'll be optimizing for maximum profit for private sectors.
- Expects a reproducible analysis. A re-runnable code that produces the same answer in a single document, easy to run code.
- First thing we should do is a summary visualization of data, but no geographical maps unless necessary for presentation.
- We should look at gas stations and model that to examine existing patterns.
- At the end, we're expected to:
    - Provide a non-technical report about our findings and the technology/software used.
    - Deliver a non-technical presentation based on our analysis, solution, next steps, and the solution's limitations.

Project Expectations

How will you know that your project was a success? List at least 3 ways that you will measure and quantify your success.

1. If it follows the patterns of current EVCS locations.
2. Constant check-in with our challenge advisor and keep him updated on our progress and take criticisms as a method of growth.
3. A comprehensive non-technical presentation is presented with solid supporting evidence and a complete conclusion.
4. Evaluation metrics applied to the solution can be explained appropriately.

Timeline and Deliverables

With your Challenge Advisor, break the semester down into smaller segments and specific steps. What tasks and outcomes do you plan to accomplish in the first few weeks, the first half of the semester, etc. List out specific steps for achieving your objectives.

Task (what will be done)	**Outcome** (expected result of task)	**Start Date**	**Deadline**
Discuss data indicators to be used	**Get a sense of what type of data we'll utilize and what data we're missing.**	**9/5**	**9/9**
Research private sector data and convert them into CSV files.	**So that data can start to be used and understood**	**9/9**	**9/11**
Research existing EVCS data and convert them into CSV files.	**So that data can start be used and understood**	**9/9**	**9/11**
Assign and understand each data document. And decide which data we should use.	**So we know what type of data we'll use to begin our data preparation**	**9/11**	**9/15**
Assign and begin cleaning, filtering data	**So that data can start to be used**	**9/15**	**9/18**

Resources

Are there any resources that might be useful to your team to get started?

- Guide sheets provided through Canvas
- Data list provided by KPMG
- Resources on how to work with ShapeFiles from our challenge advisor
- Direct communication to challenge advisor
- Communication with Studio Advisor

PART 3: DATA UNDERSTANDING

Data Structure and Source

What is the source of the data? What is the data type? (e.g. numerical, time series, text, images, etc.) Format? (e.g. tabular, nested, array, etc.) How much data is provided? Or, will you need to procure your own data?

- City of Dallas GIS Services
    - Format: shapefiles
- Texas Open Data Portal:
    - Formats: JSON, PB File, BI Dashboards, etc.
- Texas Department of Transportation
    - Data type: numerical, text
    - Formats: JSON, CSV, etc.
- My Geodata Cloud:
    - Formats: OSM, shapefile
- Zip Codes:
    - Data type: numerical, text
    - Formats: CSV
- US Department of Energy:
    - Data type: text, numerical
    - Formats: JSON, GeoJSON, XML, KML, CSV
- Procuring our own data:
    - Google

Data Understanding

In your own words, what are some of the features of the data provided to you?

- Highways in the area
- Electric fuels available / electric grid
- Number of existing stations
- Connectors between stations
- Type of disadvantaged in the area
- Existing grocery stores (geometric coordinates)
- Existing EVCS and their location coordinates

Data Preparation

What data preprocessing and/or augmentation might be required?

- Determine the specific datasets we are going to use
- Deal with missingness: determine the type of missing and decide how much we care about it and imputation methods (missing at random -> fill in the mean)
- Combine the zip code and geographical data
    - Determine geographical data organization (by zipcode, by region, overlaps, etc.)
- Converting the shape file into a feature

Data Integrity

Does your Challenge Advisor have any expectations for where data is stored? (e.g. subfolders, annotation files, etc.)

- We can use Google Drive and Google Collab but since our challenge advisor can't access them from his work laptop, we'll also share everything on Github.
- Challenge Advisor is also a collaborator to our Github Repository.
- Code in one file.

PART 4: WAYS OF WORKING

Meeting Time

Please identify a day / time to meet with your Challenge Advisor every two weeks.

- Advisor has no problem meeting every week or multiple times a week
- As we get more in-depth into the project we'll transition from meeting bi-weekly to weekly
- For any changes we'll keep our Advisor updated
- [CURRENT] Dates and Time:
    - **Thursday (4:30 PM - 5:30 PM) and Friday (10:30 AM)**

- **Alternative between Thursday and Friday BI-WEEKLY**

Ways of Working

How will you communicate with your Challenge Advisor? What norms do you agree to? (e.g. sending a meeting agenda 24 hours in advance) How will you communicate outside of your meetings every two weeks, if at all?

- Communication methods: Slack, Gmail
- Discuss next meeting time at the end of each meeting
- Send meeting agenda before/after every meeting
- Utilize when2meet/lettucemeet
- Keep ourselves updated in Github

Project Stakeholders

Are there any other stakeholders that your Challenge Advisor would like you to consider and/or speak to at any point?

No.
All stakeholders: assumed client, Dallas EV owners, environment/transport department, businesses near EVCS