# Capstone

Project management

* Application of knowledge skills tools and techniques to meet project requirements
* Helps the company: increase success, grow internal resources, manage change, grow capabilities and expertise
* General framework:
  + Start the project, understand requirements, plan it out, do it, verify it, finish
  + “Are you making progress in time? Are you making progress in cost?”

Data Science Lifecycle

* Business understanding
* Data mining
* Data cleaning
* EDA
* Feature engineering
* Modeling
* Viz

Agrograph:

* Ag tech. Provide “field-level” risk intelligence to ag banks, crop insurers and grain originators via ML. “credit score of agriculture”
* Risk score produced via data acquisition, cleaning, feature engineering, prediction using ML algorithms
* Project pitch: yield prediction with noisy limited training data
  + Cloud computing, yield prediction via ML

Databricks:

* Lakehouse platform: unify data, analytics, and AI workloads
* Homelessness:
  + City data, income, weather, housing prices, augment with your own
  + Insight extraction
    - How does homelessness differ by region, demographics, state policy, etc.
    - Which factors affect homelessness?
    - Use within db notebooks, potential blog post.
  + Python, sql

Paychex:

* Develop a model to predict which paychex clients are more likely to purchase professional employer organization(PEO). These are HR services
* Data: each row is a client, and for each client can have multiple rows for each employee. Think it is aggregated. Anonymized.
* EDA, upsell models, performance metrics of the problem, model deployment? Regular meetings/reports, code(R), final presentation

UR Electrical Team

* A model to forecast demand
* Output CI for demand each hour for the next 72 hours
* Output peak demand for the current day and the next 2 days

Mindglyph

* Retail company: OpenCV
* Image analysis to gain insight from in person customers
* ‘product affinity’ detection
* 1 Product recommendation system based on customer purchases, gaze, demographic, behavior, etc. ML modeling
* 2 ML productization using lambda architecture: database engineer task – Kafka, spark, Hadoop, etc.
* 3 Dense object detection and spatial mapping: deep learning models
* 4 Product affinity analysis using gaze tracking

MacroXstudio

* How do we generate alpha? Something unknown and useful

Wegmans

* Customer competitor lookalike model whole foods, predict a customers propensity to shop at whole foods

Paychex 2

* Correlation between unemployment, significant events, small business data and the impact on Paychex’s client base?
* What external indicators have the highest impact on Paychex’s success?
* What are the driving indicators for new sales and new losses

RTS

* Cancellation modeling

GIDS

* Admissions insights/predict admissions

1. DataBricks
2. UR utilities
3. Agrograph
4. Paychex 1
5. Wegmans
6. RTS
7. MacroXStudio Twitter
8. MacroXStudio Nightlights