# **Teacher Notes for Data-Driven Risk Estimation**

## **Motivation and Essential Understandings**

Estimates rooted in real-world data are preferable to those from models and guestimates. You can use data to test and modify already assumed probabilities of risk events and to reshape your risk mitigation plan. Rick Tison’s “Data-Driven Estimating” [article](https://www.fminet.com/fmi-quarterly/article/2015/06/data-driven-estimating/) presents data-driven decision making in the construction industry.

* How can a project manager use data to make better decisions for project planning?

## **Context and Dataset**

Students will interpret trends for risks due to project failures using data from the **PERIL** dataset, published by Tom Kendrick.

This lesson was a part of a semester-long project-based learning exercise. Students work in teams to create a project plan for building and deploying an autonomous vehicle (AV). Before this lesson, the instructor should teach the four facets of project risk management: Risk Identification, Risk Evaluation, Risk Response, Risk Control.

## **Learning Objectives**

Students will be able to:

## Explore root causes of project failures using real-world data

## Use data visualizations as a risk assessment tool

## Modify risk estimates for autonomous project using data-driven techniques

## **Data Science Concepts and Skills**

* Summary statistics
* Exploratory data analysis; Static and interactive data visualization
* Data wrangling; data dictionary

## **Students**

This lesson is for early graduate and late undergraduate students. Students should be familiar with statistical concepts, basic data visualizations, and have worked in Excel. Though visualizations are produced using Python, students will not need to perform hands-on exercises in Python or R.

## **Time to Teach this Lesson**

This lesson can be taught in 2 sessions, with time in between for students to work in their teams. For teaching online:

**First Week**: 1-hour prep, 2-hour class session; independent teamwork

**Second Week**: 2-hour class session + team presentations and follow up

## **Lesson Materials**

You will find all the lesson materials in the GenAI GitHub repository. The Jupyter notebook is not necessary to teach this lesson but is available to those who wish to teach more hands-on Data Science.

|  |  |  |
| --- | --- | --- |
| **Materials** | **File** | **Description** |
| Lecture Notes | Lecture\_doc\_Data\_Driven\_Risk\_Estimation\_2020.docx | Word document of lecture, links, tech guidance |
| Visualizations |  | Slides of Visualizations |
| Dataset |  | Cleaned PERIL dataset (in Git) |
| Data dictionary |  | Pdf of data dictionary explaining the column headings (data fields) in the datasets |
| Jupyter notebook |  | Python scripts for visualizations and predictive model in an annotated Jupyter notebook |
| Jupyter notebook pdf |  | Pdf version of annotated Jupyter notebook |
| Template | Document included in Lecture Notes | Lesson planner with links to resources |

## **Teaching Strategies**

* Warm up exercise: Present an exemplar risk assessment of your choice.
* Pose challenge questions for engagement and allow students to interpret visualizations and hypothesize. Students may have difficulty limiting inferences to within the scope of the dataset, so discuss over-hypothesizing beyond the data.
* Illustrate improving interpretation by choosing the right chart: [Data Visualization – How to Pick the Right Chart Type?](https://eazybi.com/blog/data_visualization_and_chart_types/)
* Emphasize valid correlations: [Spurious Correlations](http://www.tylervigen.com/spurious-correlations)

## **Lesson Narrative**

|  |
| --- |
| **Module 0: Pre-lesson** |

Students should be reading articles post at Tom Kendrick’s [Failure Proof Projects](http://www.failureproofprojects.com/peril2.php). Focus students on understanding the PERIL dataset and parameters that would be applicable to their AV project. Specifically, based on the readings, what parameters seem to measure a successful project?

|  |
| --- |
| **Module 1: Present Risk Events from Previous Lesson** |

1. Ask teams to present their risk events for their AV projects and discuss: Risk ID, Risk Assessment, Risk Response, and Risk Monitoring & Control. How did they calculate their risk assessment in terms of probability of occurrence and project impact? What categories for project failures would impact their risk events?

|  |
| --- |
| **Module 2: Datasets and Exploratory Data Analysis** |

1. Introduce the PERIL dataset

* Review project failure categories
* Discuss feature definitions and values; added features:
  + TRL = Technology Readiness Level
  + Cost = Cost as a measure of impact

1. Explore features and correlations of Interest

Present correlation matrix. What features (columns) are correlated?

1. Use Exploratory Data Analysis (EDA) visualizations (from Notebook visualizations) to build inferences. Ask students to hypothesize.

* The box & whisker plot displays the distribution of cost. Which risk subcategories contribute to median impact? Which contribute to the outliers?
* Explore impact by Region. Would switching the project to another region change impact?
* Explore impact by TRL. How does TRL affect impact? What can you say about TRL as a predictor of impact?
* How does the cost vary from year to year? Does it change with Region? What subcategories contribute to the low/high costs in years 2012 and 2015 for the Americas?

1. Revisit how teams assessed risk: (a) The probability of occurrence of a risk event (b) its impact on the project. Which features affect their risk estimates? How to extract probabilities from the data visualizations?

|  |
| --- |
| **Module 3: Explain Predictive Modeling** |

|  |
| --- |
| **ASSIGNMENT:** Ask the teams to use the data visualizations to construct hypotheses based on their risk events. Ask them to blog or email their modified or new risk events. Revisit how to calculate probability of occurrence. |

|  |
| --- |
| **Module 4: Close Out** |

1. Ask the teams to present their new or modified risk events based on interpreting the data visualizations from PERIL. Formal write up to be included in project plan documentation.
2. Post assessment questions
   * How might you modify your previous risk likelihood and impact assessment?
   * Do your risk events change?
   * Do the features you’ve chosen change your determination of dominance?
   * How would you have assessed your risk without using this dataset?
   * Are there other correlations, if given more time, that would affect your risk register?