**DDC Data Challenge**

The purpose of this challenge is to give you the opportunity to work through solving problems with real NYC DDC data. You are welcome to use whatever technologies, methods, or techniques you would like to apply. You are also encouraged to incorporate as many different data sets and techniques into your solution as you think are relevant. We understand that this is potentially a huge piece of analysis, and that you have limited time – while you’re prioritizing your efforts, remember that we’re less interested in some “right” answer, and more interested in seeing your thought process in action.

**Use Case – Identifying Drivers for Schedule Durations**

One of NYC DDC goals is to complete construction projects in expeditious manner:

***Identify factors that may influence project durations***

*NYC DDC manages a portfolio of projects (building projects and road and public utility infrastructure).* Construction conditions are unique for each project varying in project complexity, nature of stakeholders, management of funding, economy and environment*.*

Imagine that you are a DDC data analytics specialist and you are asked to run analysis to identify factors that influence project duration. What analysis would you run?

While this challenge is meant to be very open-ended, **your response should include the following:**

(1) A one-page (maximum) summary of the key findings/takeaways from your analysis

(2) A separate document to explain the choices you made and how you would utilize the resources (e.i. team of five or six) under your supervision to put together your analysis

This can include a description of:

• Your conceptual framework

- What problem were you trying to solve?

- What were the questions that informed how your approach?

• Putting your framework into action

- How do the summary statistics and cuts of the data you used in the analysis support your approach?

- How did you choose the models and/or techniques you used for your analysis?

- No data is perfect. How did you factor that into your approach? Does it affect how you interpreted your results?

- What other (internal or external) data would you incorporate in order to develop a more solid predictive model?

• Implementing and communicating your results

- How can the City use this information to reach its goals?

- What other information would be useful for decision-makers? (3) Relevant supporting files, which can include Excel, code, etc.

And finally – analytics is iterative! Feel free to describe any next steps and what you would have included given more time / unlimited access to data.

**Data sources:**

* Spreadsheet attached in the email

Two great sources of data are available at:

* DDC database on project schedule and cost on NYC Open Data
* NYC Open Data Portal: https://nycopendata.socrata.com/

**Please submit all of your material electronically to** [controls@ddc.nyc.gov](mailto:controls@ddc.nyc.gov%20) **five days from receipt of this email. In your email, please acknowledge that all work is your own and include the statement below:**

*With this submission, I certify that all work is my own and has not relied on consultation with others. Where information or ideas are derived from others, it has been clearly recognized and acknowledged.*

**If you have any questions regarding this challenge, please contact us through the** [[controls@ddc.nyc.gov](mailto:controls@ddc.nyc.gov%20)](mailto:modajobs@cityhall.nyc.gov) **address.**