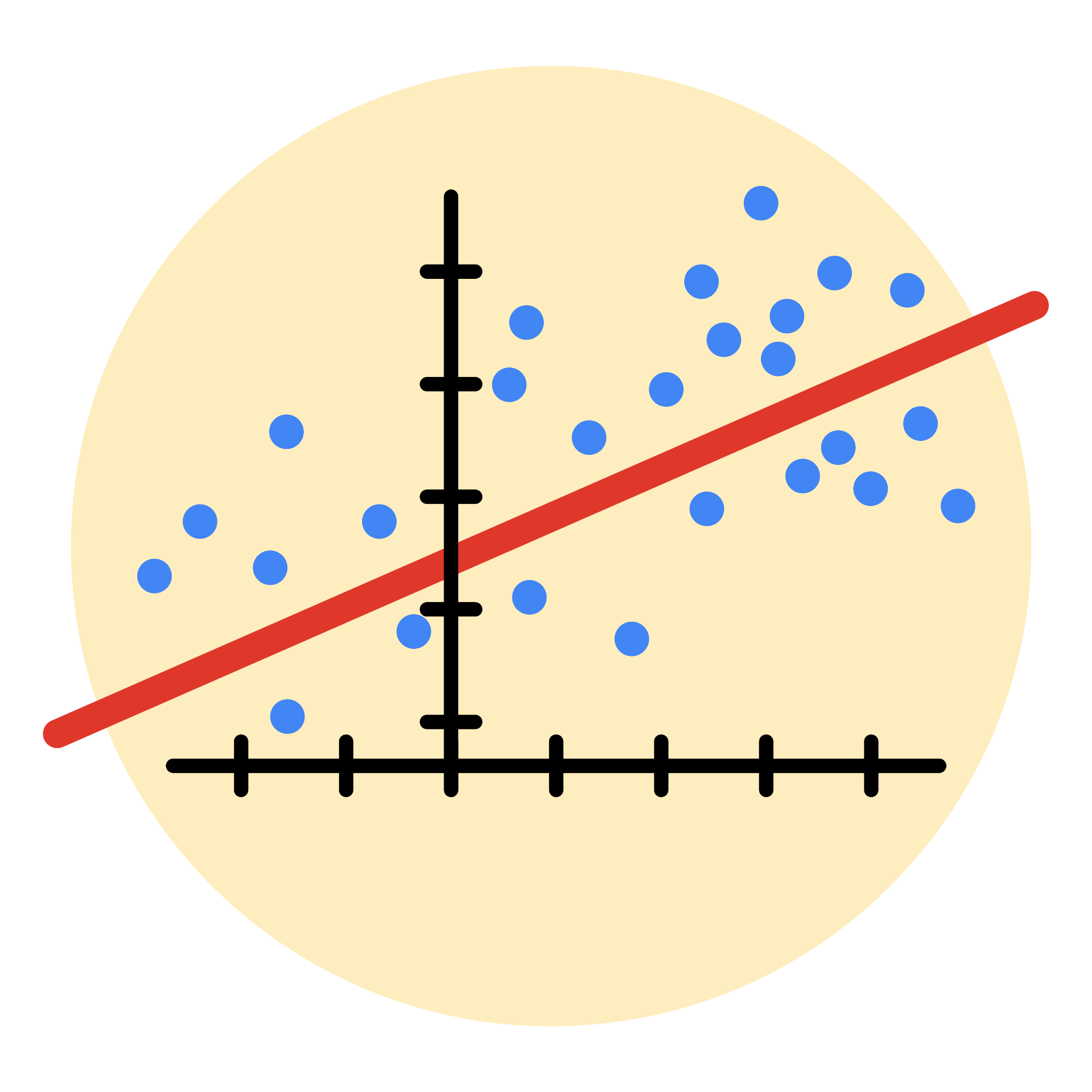
**Course Five**

# Regression Analysis: Simplifying Complex Data Relationships



# Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. As a reminder, this document is a resource that you can reference in the future, and a guide to help you consider responses and reflections posed at various points throughout projects.

# Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

* Complete the questions in the Course 5 PACE strategy document
* Answer the questions in the Jupyter notebook project file
* Build a multiple linear regression model
* Evaluate the model
* Create an executive summary for team members

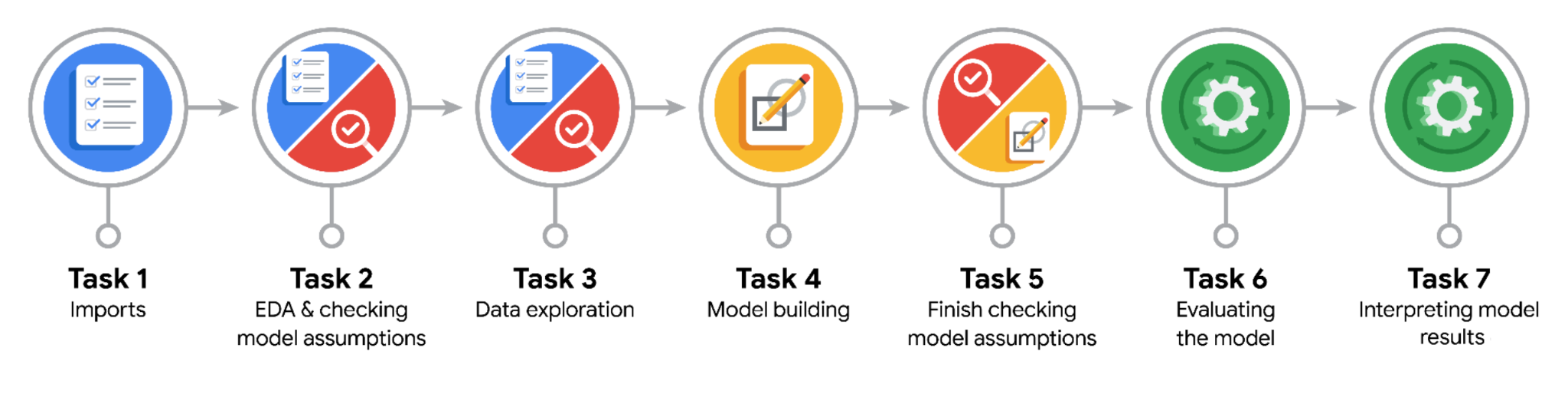
# Relevant Interview Questions

Completing the end-of-course project will empower you to respond to the following interview topics:

* Describe the steps you would take to run a regression-based analysis
* List and describe the critical [assumptions of linear regression](https://www.digitalvidya.com/blog/assumptions-of-linear-regression/)
* What is the primary difference between R2 and adjusted R2?
* How do you interpret a Q-Q plot in a linear regression model?
* What is the bias-variance tradeoff? How does it relate to building a multiple linear regression model? Consider variable selection and adjusted R2.

**Reference Guide**

This project has seven tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* Who are your external stakeholders for this project?

Mary Joanna Rodgers- Project Management Officer  
Margery Adebowale- Finance Lead, Americas  
Maika Abadi- Operations Lead

* What are you trying to solve or accomplish?

The general ideas is verified users are more likely to post opinions. That will help to predict if videos or comments from user is a claim or an opinion.

This project focus on how to predict a user is verified or not.

* What are your initial observations when you explore the data?

The initial observations is verified users are more likely to post opinions.

* What resources do you find yourself using as you complete this stage?

As email from Data Science Manager, data team will use the fictional test dataset (synthetic data) to build and evaluate a logistic regression model. It is tiktok\_dataset.csv.

**PACE: Analyze Stage**

* What are some purposes of EDA before constructing a multiple linear regression model?

Purpose of EDA before constructing model:

- Know about characteristic of data: size of data (columns x rows), how many variables and data types, descriptive statistics of every variables => give the general view about dataset and help to decide with variables (or derivative variable) with helpful for model.

- Clean data: handle with missing values, outliers

- Transform data for fitting with input requirement: transform data distribution, creating new variables that helpful to model (from existing variables)

* Do you have any ethical considerations in this stage?

I see no ethical considerations in this stage because this ‘s the fictional test dataset with no naming variable.

**PACE: Construct Stage**

* Do you notice anything odd?

video\_duration\_sec should be float type because we can use it as a numeric variable.

* Can you improve it? Is there anything you would change about the model?

Change data type of video\_duration\_sec into float.

* What resources do you find yourself using as you complete this stage?

**PACE: Execute Stage**

* What key insights emerged from your model(s)?
* What business recommendations do you propose based on the models built?
* To interpret model results, why is it important to interpret the beta coefficients?
* What potential recommendations would you make?
* Do you think your model could be improved? Why or why not? How?
* What business/organizational recommendations would you propose based on the models built?
* Given what you know about the data and the models you were using, what other questions could you address for the team?
* Do you have any ethical considerations at this stage?