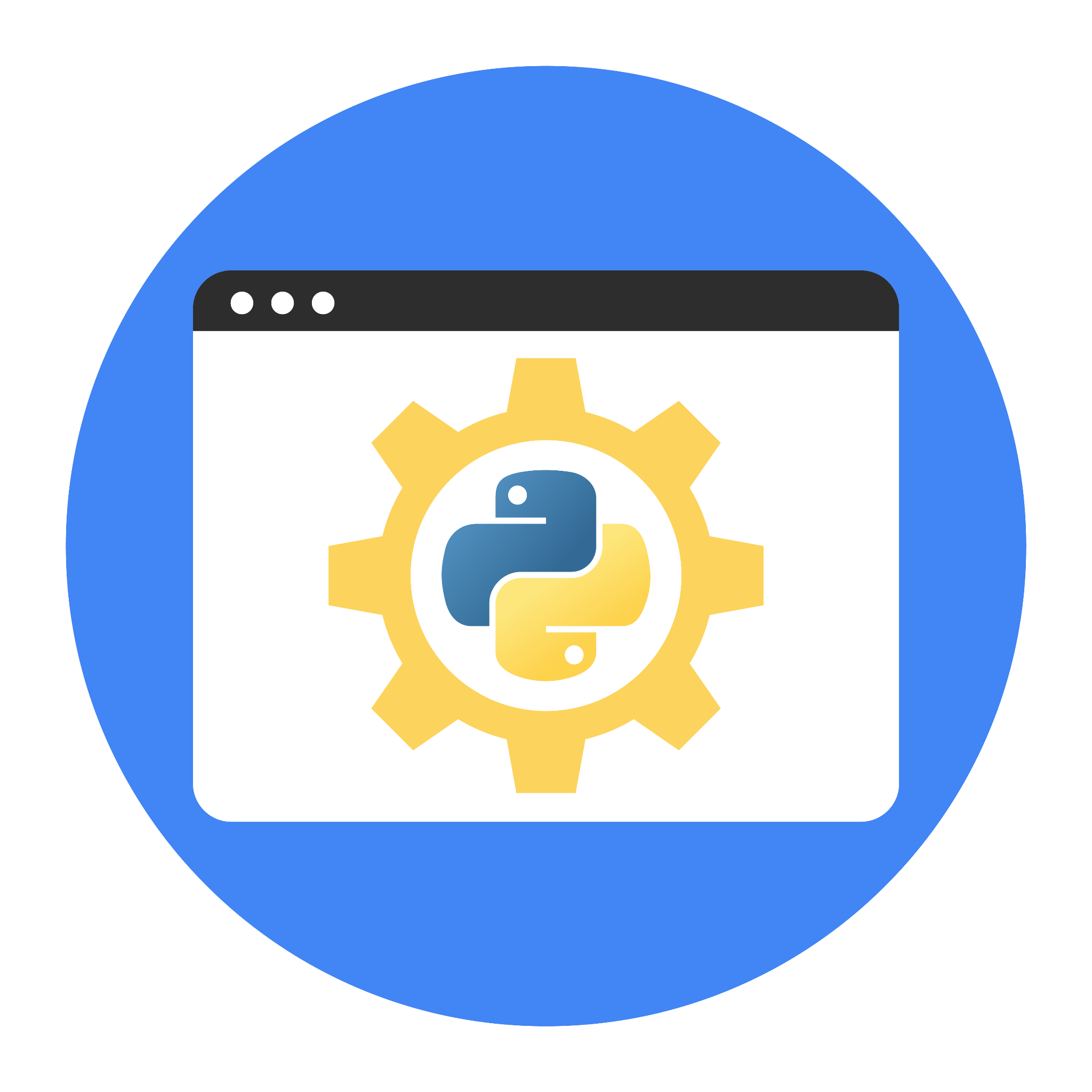
**Course Two**

# Get Started with Python



# Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future 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 2 PACE strategy document
* Answer the questions in the Jupyter notebook project file
* Complete coding prep work on project’s Jupyter notebook
* Summarize the column Dtypes
* Communicate important findings in the form of an executive summary

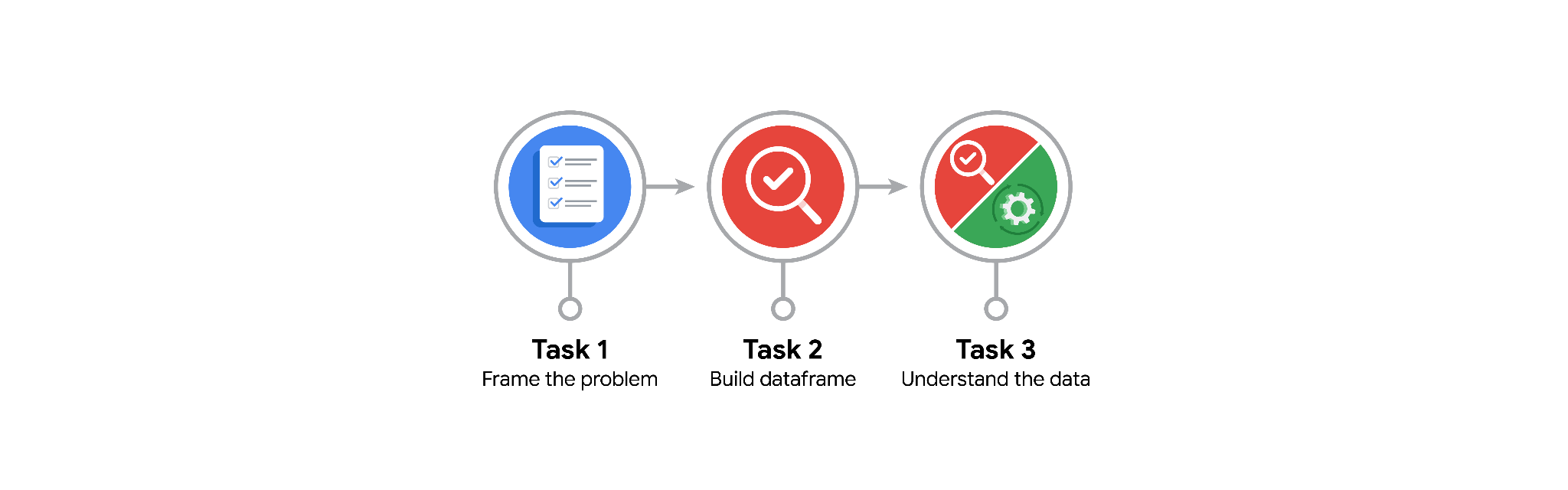
# Relevant Interview Questions

Completing the end-of-course project will help you respond these types of questions that are often asked during the interview process:

* Describe the steps you would take to clean and transform an unstructured data set.
* What specific things might you look for as part of your cleaning process?
* What are some of the outliers, anomalies, or unusual things you might look for in the data cleaning process that might impact analyses or ability to create insights?

**Reference Guide**

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



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* How can you best prepare to understand and organize the provided information?

**E-postaları dikkatlice okuyun** ve beklentileri netleştirin.

**Veri setini inceleyin**, veri türlerini (dtypes), eksik verileri ve gereksiz sütunları belirleyin.

**Veriyi anlamlı hale getirin**, gerekirse sütunları birleştirerek yeni değişkenler oluşturun.

**Bulgularınızı Jupyter not defterine ekleyin** ve net bir şekilde raporlayın.

* What follow-along and self-review codebooks will help you perform this work?

**Takip edilecek kod defterleri**: Veri keşfi (EDA), veri temizleme ve dönüşüm süreçlerini içeren örnekler.

**Kendi inceleme kod defterleri**: Veri türleri, eksik veriler ve sütun analizlerini kontrol etmek için rehber niteliğinde not defterleri.

* What are some additional activities a resourceful learner would perform before starting to code?

**Proje gereksinimlerini anlama** – E-postaları, dokümanları ve veri seti açıklamalarını dikkatlice inceleyerek neyin gerekli olduğunu belirler.

**Benzer projeleri araştırma** – TLC verisiyle ilgili önceki çalışmalara veya benzer veri analizlerine göz atar.

**Veri yapısını analiz etme** – Veri setinin önizlemesini yaparak sütun isimlerini, veri türlerini ve eksik değerleri anlamaya çalışır.

**Önemli değişkenleri belirleme** – Hangi sütunların analiz için kritik olduğunu değerlendirir.

**Kodlama stratejisi oluşturma** – Kullanılacak araçları, paketleri ve adımları planlayarak sistemli bir yol haritası çizer.

**PACE: Analyze Stage**

* Will the available information be sufficient to achieve the goal based on your intuition and the analysis of the variables?

Mevcut bilginin hedefe ulaşmak için yeterli olup olmadığı, veri setinin kapsamına ve kalitesine bağlıdır. Eğer veri türleri, eksik değerler ve ilgili değişkenler analiz edildikten sonra gerekli bilgiler eksiksizse, hedefe ulaşmak mümkün olacaktır. Ancak, veri setinde eksiklikler veya yetersiz değişkenler varsa, ek veri kaynağına veya ön işleme adımlarına ihtiyaç duyulabilir. Bu nedenle, ön analiz yaparak veri setinin proje gereksinimlerini karşılayıp karşılamadığını değerlendirmek önemlidir.

* How would you build summary dataframe statistics and assess the min and max range of the data?

**Veriyi İnceleyin:** Veri setini yükleyerek sütun isimlerini, veri türlerini ve genel yapısını kontrol edin.

**Özet İstatistikleri Çıkarın:** Veri setindeki sayısal sütunların ortalama, medyan, standart sapma, minimum ve maksimum gibi temel istatistiklerini inceleyin.

**Minimum ve Maksimum Değerleri Analiz Edin:** Hangi sütunların en düşük ve en yüksek değerlere sahip olduğunu belirleyerek anormal veya beklenmeyen değerler olup olmadığını kontrol edin.

**Veri Dağılımını Görselleştirin:** Histogram veya kutu grafikleri kullanarak değişkenlerin dağılımını analiz edin. Bu, uç değerleri (outlier) ve veri aralığını anlamaya yardımcı olur.

**Eksik veya Hatalı Verileri Kontrol Edin:** Eksik verileri ve beklenmeyen negatif ya da aşırı uç değerleri tespit ederek veri temizleme gereksinimini değerlendirin.

* Do the averages of any of the data variables look unusual? Can you describe the interval data?

**Aralık (interval) verileri:** Sayısal değişkenlerin minimum ve maksimum değerlerini analiz ederek veri dağılımını anlamak mümkündür. Eğer ortalama, beklenenden çok yüksek veya düşükse, uç değerler veya hatalı veriler olabilir. Bu durumda veri temizleme gerekebilir.

**PACE: Construct Stage**

**Note**: The Construct stage does not apply to this workflow. The PACE framework can be adapted to fit the specific requirements of any project.

**PACE: Execute Stage**

* Given your current knowledge of the data, what would you initially recommend to your manager to investigate further prior to performing exploratory data analysis?

**Eksik ve tutarsız veriler:** Hangi sütunlarda eksik veya hatalı değerler olduğunu belirlemek.

**İlgisiz sütunların filtrelenmesi:** Analiz için gereksiz değişkenleri ayıklamak.

**Aykırı değerler:** Ortalama ve dağılım analizleri ile uç değerleri tespit etmek.

**Değişken dönüşümleri:** Gerekirse yeni anlamlı değişkenler oluşturmak veya mevcutları dönüştürmek.

* What data initially presents as containing anomalies?

**Aykırı değerler:** Minimum ve maksimum değerleri aşırı yüksek veya düşük olan sütunlar.

**Eksik veya tutarsız veriler:** Boş hücreler, eksik değerler veya mantıksız girişler (örneğin, negatif ücret değerleri).

**Kategorik değişkenlerde hatalar:** Yanlış yazılmış veya beklenmeyen kategori değerleri.

**Zaman verilerinde tutarsızlık:** Mantıksız tarih ve saat girişleri (örneğin, bitiş süresinden önce başlayan yolculuklar).

* What additional types of data could strengthen this dataset?

**Hava durumu verileri:** Yolculuk sırasında hava koşullarının etkisini analiz etmek için.

**Trafik yoğunluğu verileri:** Seyahat süresi ve ücretlendirme üzerindeki etkileri anlamak için.

**Sürücü ve araç bilgileri:** Deneyim, çalışma saatleri veya araç türüne göre analiz yapmak için.

**Özel etkinlik verileri:** Konserler, spor etkinlikleri gibi faktörlerin yolculuk taleplerine etkisini incelemek için.

**Müşteri değerlendirme verileri:** Yolculuk memnuniyeti ve kalite analizi için.