**Exploratory Data Analysis (EDA) Project Guide – Standing Out**

**Step-by-Step Process with Standout Strategies**

**1. Define Your Objective with Industry Relevance**

* Clarify the problem and purpose statement.
* Example: "Analyzing clinical trial efficiency to identify factors impacting completion times."
* **Standout Tip**: Research how Málaga-based companies (like Bioherent) approach clinical trials and frame your problem accordingly.

**2. Data Cleaning & Preprocessing**

* **Key Tasks**:
  + Handle missing values with imputation or removal.
  + Convert data types (dates, categories, numerics).
  + Remove duplicates and standardize formats.
  + Identify and handle outliers with domain-specific reasoning.
  + Standardization & Normalization: For better visualization and analysis.
* **Standout Tip**: Document your cleaning process clearly to showcase data handling skills, emphasizing how these decisions impact business insights.

**3. Perform Exploratory Data Analysis (EDA)**

* **First Look**:
  + Use .info(), .describe(), and .head() to understand structure.
  + Check for imbalances or skewed distributions.
  + Perform correlation analysis to detect relationships.
  + Feature Engineering: Create meaningful new variables if necessary.
* **Standout Tip**: Highlight domain-specific findings in your EDA. Example: "Most delays in clinical trials occur due to slow patient recruitment—what factors contribute to this?"

**5. Conduct Statistical & Advanced Analysis**

* **Descriptive Statistics**: Mean, median, standard deviation.
* **Correlations:** Find relationships between variables using heatmaps.
* **Hypothesis Testing**: Chi-square, ANOVA, or regression models for deeper insights.
* **Standout Tip**: Tie insights to real-world decisions. Example: "Our statistical tests confirm that reducing dropout rates by 5% accelerates trial completion by 2 months."

**6. Data Visualization & Communication**

* **Visuals to Use**:
  + Univariate (histograms, bar charts).
  + Bivariate (scatter plots, boxplots, correlation heatmaps).
  + Multivariate (PCA, clustering, geospatial maps if location-based).
* **Standout Tip**: Instead of generic charts, tailor visualizations to **Málaga-based biotech needs**. Example: "Heatmap of clinical trials with longest delays—highlighting optimization opportunities."
* **Tableau vs. Power BI Decision**: You can decide later based on employer preferences.

**7. Extract Business Insights**

* **Convert Data to Actionable Insights**:
  + What are the biggest bottlenecks in clinical trials?
  + Are there location-specific trends affecting efficiency?
* **Standout Tip**: Relate insights to Málaga-based companies and broader biotech trends. Example: "Our findings suggest Bioherent can streamline drug allergy testing through predictive analytics."

**8. Document & Present Findings**

* **Formats to Use**:
  + Jupyter Notebook with Markdown.
  + Well-structured GitHub README.
  + Dashboard (Tableau or Power BI).
* **Standout Tip**: Use storytelling to make insights compelling: "Imagine reducing clinical trial delays by 3 months—how much would that save companies in Málaga?"

**9. Make Your Project Stand Out**

• Personalization: Tailor insights for Málaga companies.

• Storytelling: Present findings narratively.

• Comparative Analysis: Benchmark against other studies.

• Interactive Elements: Use dashboards for engagement.

**10. Share Your Work & Get Feedback**

* Publish on **GitHub, Kaggle, or Medium**.
* Share insights with **healthcare professionals on LinkedIn**.
* Seek feedback from industry professionals, iterate, and improve.
* **Standout Tip**: Write a LinkedIn post about how **this project can benefit Málaga’s biotech industry**.

By following these steps with standout strategies, you ensure your project is **not just another EDA**, but a portfolio piece that aligns with industry needs in Málaga. Let me know if you need further guidance!