**Тест по EDA - 1**

**1.** Что является основной целью Exploratory Data Analysis?  
A) Автоматическое обучение модели  
B) Проверка гипотез и подготовка данных к моделированию  
C) Визуализация данных и выявление скрытых закономерностей  
D) Оптимизация гиперпараметров

**2.** Какой график лучше всего использовать для анализа распределения числовой переменной?  
A) Boxplot  
B) Histogram  
C) Scatter plot  
D) Heatmap

**3** Что делает метод df.describe() в Pandas?  
A) Показывает информацию о типах данных и количестве пропусков  
B) Возвращает основные статистические характеристики числовых (и иногда категориальных) переменных  
C) Строит гистограммы  
D) Считает корреляцию между признаками

**4.** Какой из методов используется для выявления пропусков в данных?  
A) df.describe()  
B) df.isnull().sum()  
C) df.corr()  
D) df.groupby()

**5.** Что показывает heatmap на основе корреляционной матрицы?  
A) Линейную зависимость между переменными  
B) Нелинейные зависимости  
C) Только выбросы  
D) Распределение категориальных переменных

**6.** Что делает метод df.info() в Pandas?  
A) Возвращает основные статистические характеристики признаков  
B) Показывает количество строк, столбцов, типы данных и количество пропусков  
C) Строит графики распределения переменных  
D) Считает стандартное отклонение и медиану

**7.** Для чего используют scatter plot?  
A) Чтобы сравнить распределения нескольких категорий  
B) Чтобы выявить взаимосвязь между двумя числовыми переменными  
C) Чтобы проверить наличие пропусков  
D) Чтобы показать медиану

**8.** Что означает высокая положительная корреляция (например, 0.9) между двумя переменными?  
A) Переменные независимы  
B) Между ними сильная линейная связь  
C) Они обязательно имеют причинно-следственную зависимость  
D) Данные некорректны

**9.** Какие методы можно использовать для уменьшения влияния выбросов?  
A) Удаление или замена медианой/средним  
B) PCA  
C) Label encoding  
D) K-means clustering

**10.** Что обычно делают в начале EDA?  
A) Запускают сложные модели  
B) Проверяют данные на пропуски, типы переменных и размер  
C) Делают feature engineering  
D) Строят только heatmap