**Amplity Health – NLP Data Analyst Assessment**

**Introduction**

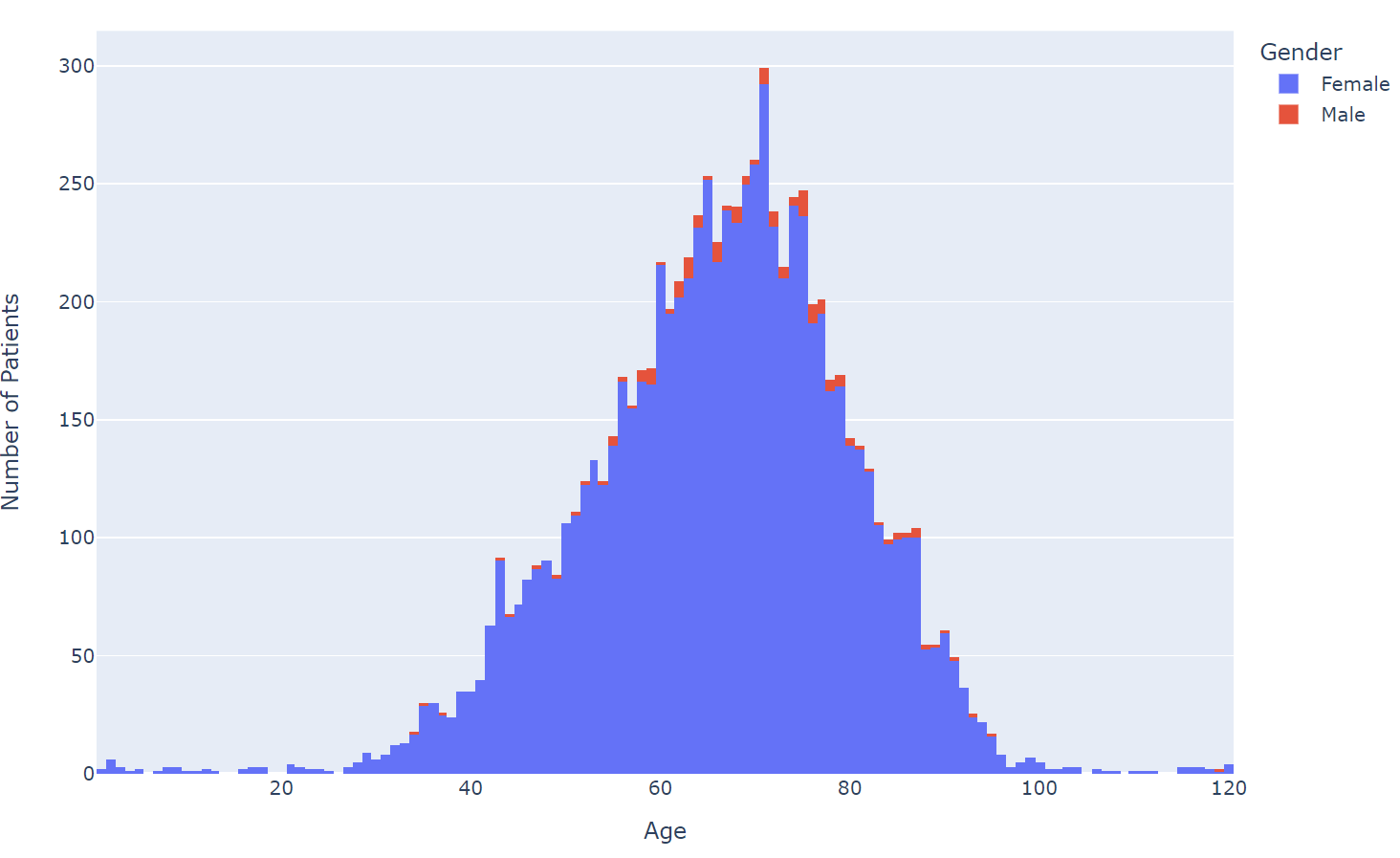
You get a dataset that includes demographic information about patients that are receiving two types of treatments. The project requires a simple overall analysis on the number of patients that corresponds to different categories. For this task, the only requirement is that you need to use python as main language.

**Dataset**

This dataset has a numerical identification for each patient (PatientID), age, gender, ethnicity, state, treatment, pain level (on scale of 1-10), source of the information (Source) and number of documents for that patient (#Docs). **\*NOTE: Despite the PatientID being a 16+ digit number, it should be loaded as data type = text to provide accurate patient counts.**

**Tasks**

**NOTE: For all of the below analysis, it should be pivoted on the PatientID, not rows. There may be more than one row per patientID. IMPORT PATIENTID AS TEXT!**

1. Obtain and print an overall age analysis (mean, min, max, median) for each treatment. For each patient, keep only the highest age encountered.
2. Plot the distribution of Age and Gender by PatientID using Plotly package. The style of the distribution should look similar to this:
   1. 
3. Obtain an overall pain level analysis (mean, min, max, median) for each treatment type, print the raw values and generate box plots for each treatment type. For individual values represented as a range, keep the higher value. For each PatientID, keep the highest value. **NOTE: This should be calculated in terms of # of patients.**
4. **NLP test**: There is a column “Verbatim” that contains raw text. Mine this information and present the following insights:
   1. Pain location – create a bar chart showing the top 5 body-parts that experience pain (a list of body-parts is included below for reference)
   2. Pain Description - create a bar chart showing the top 5 types of pain (a list of possible pain descriptors is below)

**Delivery**

The delivery for this task should be:

* HTML File: A Jupyter notebook exported as HTML that you used to resolve this task and all the outputs obtained. Include any explanations and thought processes as markup blocks above the associated code block.

**Honors Code**

By submitting this exercise, I agree that I will submit my own work and only my own work. I will not submit the work of any other person nor AI-generated work. I will not engage in any activity that would falsify or misrepresent my results or the results of others.

Lists:

**Body-parts:**

* Lower back
* upper back
* shoulder
* neck
* hip
* fingers
* ankle
* arm
* abdomen
* elbow
* wrists
* legs
* joint
* wrist
* foot
* hand
* pelvis
* feet
* knee
* hands
* buttocks
* thighs

**Pain Descriptors:**

* aching
* sharp
* dull
* stiffness
* intermittent
* persistent
* burning
* throbbing
* localized
* tightness
* stabbing
* radiating