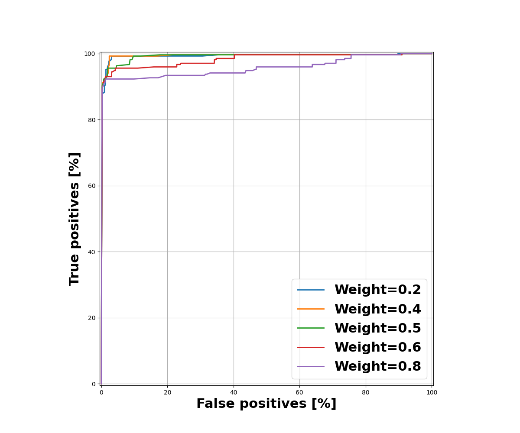
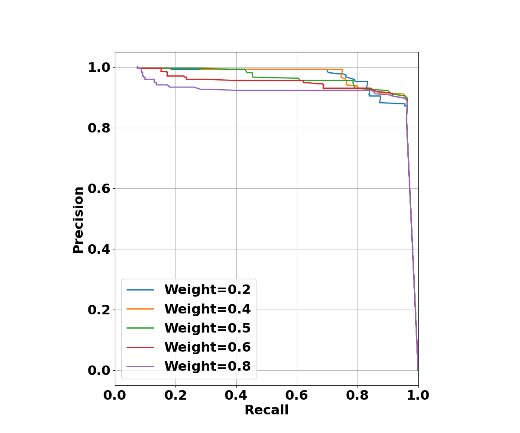
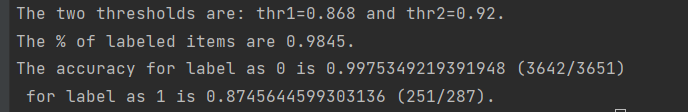
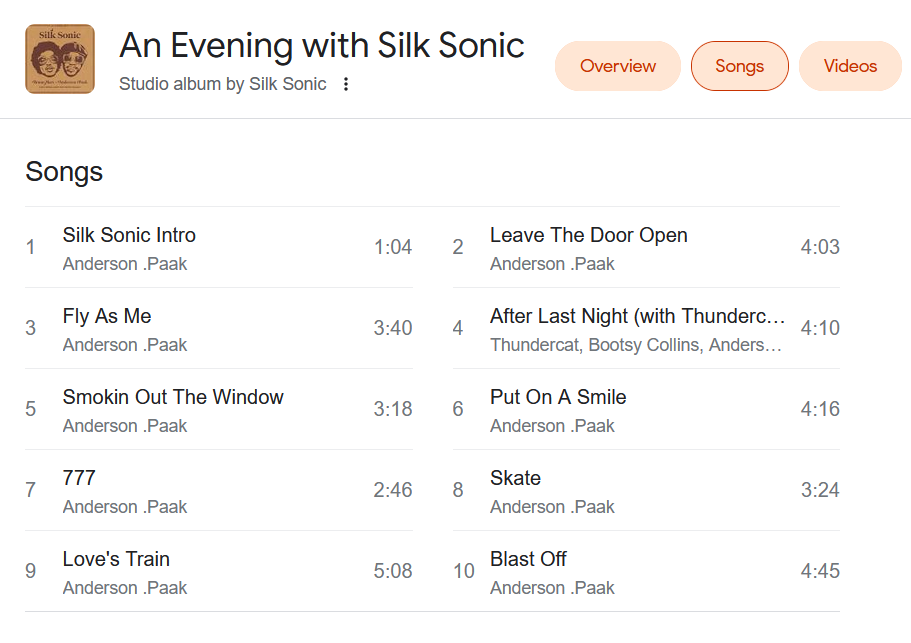
**Result analysis of the Current Progress**

1. **PRC and ROC analysis**: (here takes 1000-5000 data for example)  
   
   1. **Goal1：**determine the weight percentage of evaluation function
      1. Evaluation function is to combine the artist fuzzy score and track fuzzy score with different weights.
      2. **Eval\_score = w\*artistFuzzyScore + (1-w)\*trackFuzzyScore**
      3. Here is to determine the weight **w** for the artist fuzzy score using the Prc and Roc curves.
      4. Area under the curve (AUC) score is applied to determine which weight **w** is better. Typically, the Prc and Roc will reach the same conclusion.
   2. **Goal2:** set 2 the proper **threshold** to categorize all the data into 3 parts:
      1. Label = 0: 0 <= eval\_score <= threshold1
      2. Label = uncertain: threshold1<eval\_score<threshold2
      3. Label = 1: threshold2 <= eval\_score <= 1
      4. \*The **current thresholds (empirical)** are set by:
2. Threshold1: precision = 0.9
3. Threshold2: recall = 0.97
4. **Evaluate result**：
   1. Console outputs:  
      
      1. The **percentage of data that are labeled as 0 or 1** are reported
      2. The **accuracy of label=0 and label=1** are both reported.
   2. In addition, the categorized data are summarized in three tables containing the data with label=0, label=uncertain, and label=1.
   3. In order to facilitate the error analysis, two more tables are generated：
      1. Incorrect label for label=0 (false negative)
      2. Incorrect label for label=1 (false positive)
   4. The following section summarizes all the problems when doing the fuzzy wuzzy and manual entity maching.
5. **Error analysis summary (for 0-8000 data)**:

|  |  |  |  |
| --- | --- | --- | --- |
| Campaign name | Artist name | Track name | Comment |
| WrIe - Silk Sonic - Silk Sonic 鈥?An Evening with Silk Sonic (Album) - P1 - R - 0062L00000WCVqnQAH [Ireland] | Bruno Mars, Anderson .Paak, Silk Sonic | Smoking Out The Window | Album vs Track name |

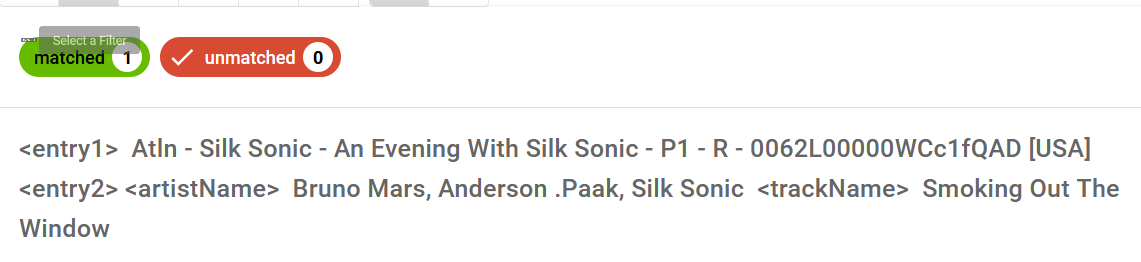
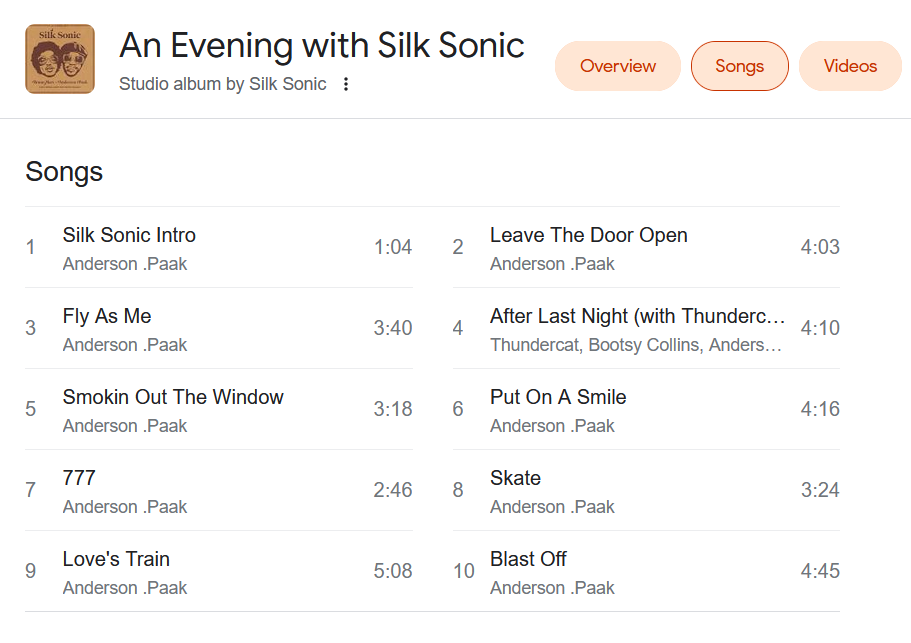
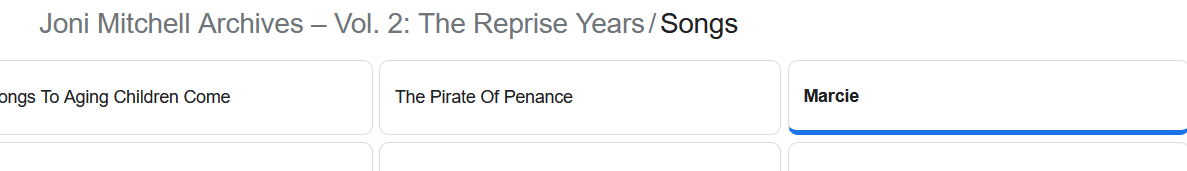
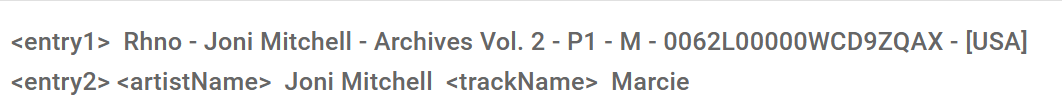
* 1. **Issue: track name vs album name**
     1. Campaign name here contains the album name, not the track name.
     2. From the search below, the album name is matching the artist name and the track name  
        
     3. \*This situation may have happened to many other inputs.
     4. Current handle: assign these inputs as Not Match, until the business people respond.

|  |  |  |  |
| --- | --- | --- | --- |
| Campaign name | Artist name | Track name | Comment |
| [A] WrNa - Warner Music Nashville - November YouTube Campaigns - P1 - R - 0062L00000WCWyiQAH | gnash | us | both artist and track names accidently match the content in campaign name |

* 1. **False Positive**:  
     1. As shown here, the artist and track names make no sense, but are both short and match in the campaign name.
     2. This suggests that **fuzzy scores with both 100 points are still not good enough to give label = 1.**
  2. **False Negative**:

|  |  |  |  |
| --- | --- | --- | --- |
| Campaign name | Artist name | Track name | Comment |
| Atln - Jasmine Thompson - Love Is Just A Word ft. Calum Scott - P1 - R - 0062L00000WCTxSQAX | Jasmine Thompson and Calum Scott | love is just a word | Artist name contains 2 musicians |
| PLUMITAS | LIKE | Plumas | Like | Possible a typo |

* + 1. Two cases:
       1. The first case is due to 2 musician names appear in the “artist name” entry, but “campaign name” only contains 1 musician name
       2. The second case is possibly due to a typo.

1. Summary of issue and limitation of fuzzy wuzzy:
   1. **Issue**: **some input in CampaignName is the album name, instead of song track name**
      1. The input is:  
         
      2. Search for the info:  
         
         1. The trackName “Smoking Out The Window” is a song in the album “An Evening with Silk Sonic”
      3. Not sure whether to label as 1 or 0
      4. Another example:  
         
   2. Fuzzy matching generates low score for **artistName contains 2 musicians while the campaign name only contains 1 musician**
   3. Seems like a **typo**:  
      
      * 1. So I manually gave the label as 1.
   4. The **artistName, trackName are too simple, and somehow both are overlapped within the campaignName**, such as:  
      
      1. **\*fuzzy score = 100 for both artistName and trackName may not be a good indication to label.**
   5. Not sure if matched:  
      
      1. Christmas is not a song name from Sia