Dataframes produced from lentries files using the Create **Dataframes Script**

Missing First Name <u>Dataframes</u>

df_year.csv(CSV files in /missing_last_name)

Missing first name dataframes ranging from 1908-1918

Missing Last Name <u>Dataframes</u>

df_year.csv(CSV files in /missing_last_name)

Missing last name dataframes ranging from 1908-1918

Missing Format Dataframes

df_year.csv(CSV files in /missing_format)

Missing format dataframes ranging from 1908-1918

Missing Price **Dataframes**

df_year.csv(CSV files in /missing_price)

Missing price dataframes ranging from 1908-1918

Missing Publisher **Dataframes**

df_year.csv(CSV files in /missing_publisher)

Missing publisher dataframes ranging from 1908-1918

Missing Title Dataframes

df_year.csv(CSV files in /missing_title)

Missing title dataframes ranging from 1908-1918

Create Entries Script

file in /scripts)

reads catalogue text files, spots and splits entries, produces measures for

Clean Entries

Entries extracted

from /princeton_years files

using the Create

Entries Script

Full Entries

entries_year.txt(text files

in /full_entries)

All entries

ranging from 1908-1918

Line Mid

entries_year.txt(text files

in /line_mid_entries)

Entries containing date

values in the middle of the

entries ranging from

1908-1918

entries_year.txt(CSV files in /clean_entries)

Clean (non front truncated and non medial line) entries ranging from 1908-1918

Front Trunc Entries

entries_year.csv(CSV files in /front_trunc_entries)

> Front Truncated Entries ranging from 1908-1918

Entries Measures

entries_measures_ year.txt(text files in /entries_measures)

Measures for different entry categorizations ranging from 1908-1918

Clean Dataframes

Create Dataframes Script

create_

dataframes.py(python file

in /scripts)

reads entry CSV files,

assigns/spots bibliographic

information to produce

dataframes, categorizes

entries, and produces

measures for categorizations df_year.csv(CSV files in /clean_dataframe)

Clean ("mostly" non-null) dataframes ranging from 1908-1918

Full Dataframes

df_year.csv(CSV files in /full_dataframe)

Full dataframes ranging from 1908-1918

Dataframe Measures

df_measures_ year.csv(CSV files in /dataframe_measures)

Measures for different dataframe categorizations ranging from 1908-1918

HathiTrust **Princeton Published Optical Character** Recognition **Transcriptions for the English Catalogue of Books**

ECB Files

ecb_year.txt(text files in /princeton_years)

ranging from 1900-1920 and 1922

create_entries.py(python

categorizes entries, and categorizations