

01b_FO

May 26, 2024

1 Convert Historical Data (txt) from Osborne & Frey to .csv

source: Osborne, M. A., & Frey, C. B. (2013). The future of employment: How susceptible are jobs to computerisation?

converted to txt with [snagit](#)

1.1 Content

1. Import TXT
2. Convert to CSV

```
[1]: # imports
import pandas as pd
import re
```

Import TXT

```
[2]: # import txt file
df = pd.read_fwf('files/F0.txt')
df.head(5)
```

```
[2]:          Rank,Probability,Label,code,Occupation
0          1. 0.0028 29-1125 Recreational Therapists
1  2. 0.003 49-1011 First-Line Supervisors of Mec...
2    3. 0.003 11-9161 Emergency Management Directors
3  4. 0.0031 21-1023 Mental Health and Substance ...
4          5. 0.0033 29-1181 Audiologists
```

Convert to CSV

```
[3]: # create csv out of txt

# Create an empty list to store the rows
rows = []
header = ','.join(df.columns.astype(str))
rows.append(header)

# Iterate through each row
for index, row in df.iterrows():
```

```

# Convert the row to a string
row_str = ' '.join(row.astype(str))
# Replace the first space with a comma
row_str = row_str.replace(' ', ',', 1)
# Check if there's a single 0 or 1 after the next space (if job is marked
↳as automatable or not)
if re.search(r' [01] ', row_str):
    # If true, replace the space with a comma
    row_str = re.sub(r' ([01]) ', r',\1,', row_str)
else:
    # If false, add "na" in between
    row_str = re.sub(r' ', ',na,', row_str, 1)

# remove last space
row_str = row_str.replace(' ', ',', 1)

# Remove all commas after the 4th comma (job description has commas in it,
↳which messes up the csv file)
parts = row_str.split(',')
row_str = ','.join(parts[:5]) + ','.join(parts[4:])
# Add the row to the list

rows.append(row_str)

```

```

[4]: # Open a file in write mode
with open('files/F0.csv', 'w') as file:
    # Write each row to the file
    for row in rows:
        file.write(row + '\n')

```