**Lab 4 Submission**

**Data Wrangler Script and Screenshot: CMSC**

from wrangler import dw

import sys

if(len(sys.argv) < 3):

sys.exit('Error: Please include an input and output file. Example python script.py input.csv output.csv')

w = dw.DataWrangler()

# Split data repeatedly on newline into rows

w.add(dw.Split(column=["data"],

table=0,

status="active",

drop=True,

result="row",

update=False,

insert\_position="right",

row=None,

on="\n",

before=None,

after=None,

ignore\_between=None,

which=1,

max=0,

positions=None,

quote\_character=None))

# Extract from data on 'CMSC any number '

w.add(dw.Extract(column=["data"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on="CMSC\\d+",

before=None,

after=None,

ignore\_between=None,

which=1,

max=1,

positions=None))

# Fill extract with values from above

w.add(dw.Fill(column=["extract"],

table=0,

status="active",

drop=False,

direction="down",

method="copy",

row=None))

# Wrap rows where data starts with '0'

w.add(dw.Wrap(column=[],

table=0,

status="active",

drop=False,

row=dw.Row(column=[],

table=0,

status="active",

drop=False,

conditions=[dw.StartsWith(column=[],

table=0,

status="active",

drop=False,

lcol="data",

value="0",

op\_str="starts with")])))

# Delete row 1

w.add(dw.Filter(column=[],

table=0,

status="active",

drop=False,

row=dw.Row(column=[],

table=0,

status="active",

drop=False,

conditions=[dw.RowIndex(column=[],

table=0,

status="active",

drop=False,

indices=[0])])))

# Drop wrap3

w.add(dw.Drop(column=["wrap3"],

table=0,

status="active",

drop=True))

# Drop wrap13, wrap12, wrap11

w.add(dw.Drop(column=["wrap13","wrap12","wrap11"],

table=0,

status="active",

drop=True))

# Drop wrap5, wrap7, wrap9

w.add(dw.Drop(column=["wrap5","wrap7","wrap9"],

table=0,

status="active",

drop=True))

# Extract from wrap4 between ': ' and ','

w.add(dw.Extract(column=["wrap4"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=",",

after=": ",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Extract from wrap4 between 'Open: ' and ','

w.add(dw.Extract(column=["wrap4"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=",",

after="Open: ",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Extract from wrap4 between 'Waitlist: ' and ')'

w.add(dw.Extract(column=["wrap4"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before="\\)",

after="Waitlist: ",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Drop wrap4

w.add(dw.Drop(column=["wrap4"],

table=0,

status="active",

drop=True))

# Extract from wrap6 before ' '

w.add(dw.Extract(column=["wrap6"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=" ",

after=None,

ignore\_between=None,

which=1,

max=1,

positions=None))

# Extract from wrap6 after ' '

w.add(dw.Extract(column=["wrap6"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=None,

after=" ",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Drop wrap6

w.add(dw.Drop(column=["wrap6"],

table=0,

status="active",

drop=True))

# Extract from wrap8 before ' '

w.add(dw.Extract(column=["wrap8"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=" ",

after=None,

ignore\_between=None,

which=1,

max=1,

positions=None))

# Extract from wrap8 after ' '

w.add(dw.Extract(column=["wrap8"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=None,

after=" ",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Drop wrap8

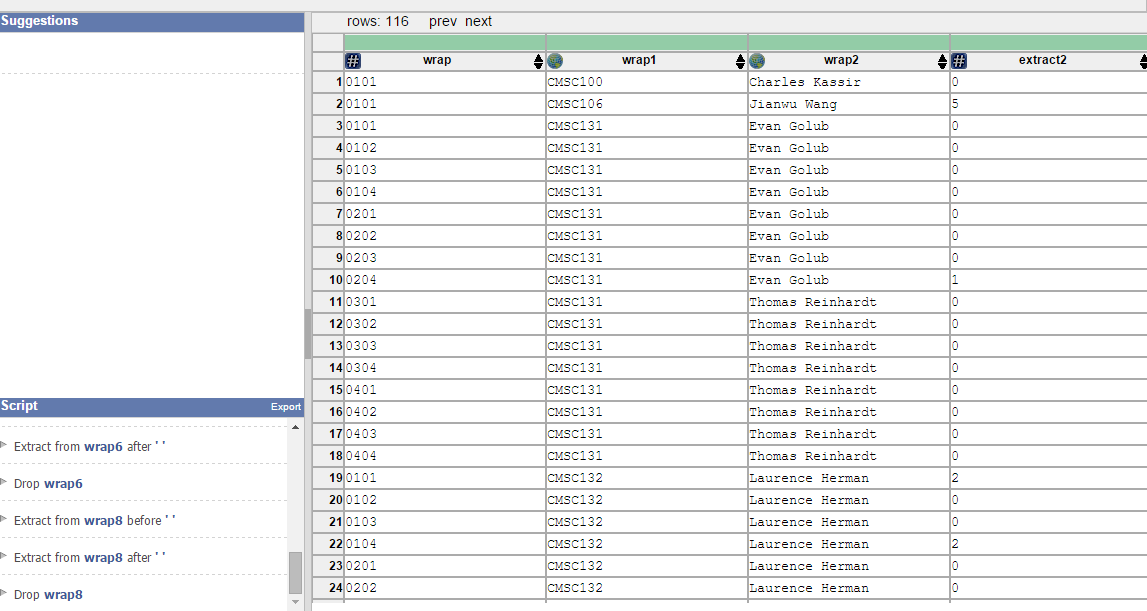
w.add(dw.Drop(column=["wrap8"],

table=0,

status="active",

drop=True))

w.apply\_to\_file(sys.argv[1]).print\_csv(sys.argv[2])



**Data Wrangler Script and Screenshot: World Cup 1**

**from wrangler import dw**

**import sys**

**if(len(sys.argv) < 3):**

**sys.exit('Error: Please include an input and output file. Example python script.py input.csv output.csv')**

**w = dw.DataWrangler()**

**# Split data repeatedly on newline into rows**

**w.add(dw.Split(column=["data"],**

**table=0,**

**status="active",**

**drop=True,**

**result="row",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on="\n",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=0,**

**positions=None,**

**quote\_character=None))**

**# Delete row 1**

**w.add(dw.Filter(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**row=dw.Row(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**conditions=[dw.RowIndex(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**indices=[0])])))**

**# Wrap rows where data = '|-'**

**w.add(dw.Wrap(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**row=dw.Row(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**conditions=[dw.Eq(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**lcol="data",**

**value="|-",**

**op\_str="=")])))**

**# Extract from wrap1 on ' any uppercase word '**

**w.add(dw.Extract(column=["wrap1"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on="[A-Z]+",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Drop wrap1**

**w.add(dw.Drop(column=["wrap1"],**

**table=0,**

**status="active",**

**drop=True))**

**# Drop wrap**

**w.add(dw.Drop(column=["wrap"],**

**table=0,**

**status="active",**

**drop=True))**

**# Set wrap2 name to 1**

**w.add(dw.SetName(column=["wrap2"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["1"],**

**header\_row=None))**

**# Set wrap3 name to 2**

**w.add(dw.SetName(column=["wrap3"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["2"],**

**header\_row=None))**

**# Set wrap4 name to 3**

**w.add(dw.SetName(column=["wrap4"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["3"],**

**header\_row=None))**

**# Set wrap5 name to 4**

**w.add(dw.SetName(column=["wrap5"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["4"],**

**header\_row=None))**

**# Drop wrap6**

**w.add(dw.Drop(column=["wrap6"],**

**table=0,**

**status="active",**

**drop=True))**

**# Drop wrap7**

**w.add(dw.Drop(column=["wrap7"],**

**table=0,**

**status="active",**

**drop=True))**

**# Fold 1, 2, 3, 4 using header as a key**

**w.add(dw.Fold(column=["\_1","\_2","\_3","\_4"],**

**table=0,**

**status="active",**

**drop=False,**

**keys=[-1]))**

**# Split value repeatedly on ',' into rows**

**w.add(dw.Split(column=["value"],**

**table=0,**

**status="active",**

**drop=True,**

**result="row",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on=",",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max="0",**

**positions=None,**

**quote\_character=None))**

**# Extract from value between ' any word any word |' and ']'**

**w.add(dw.Extract(column=["value"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on=".\*",**

**before="]",**

**after="[a-zA-Z]+[a-zA-Z]+\\|",**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Drop value**

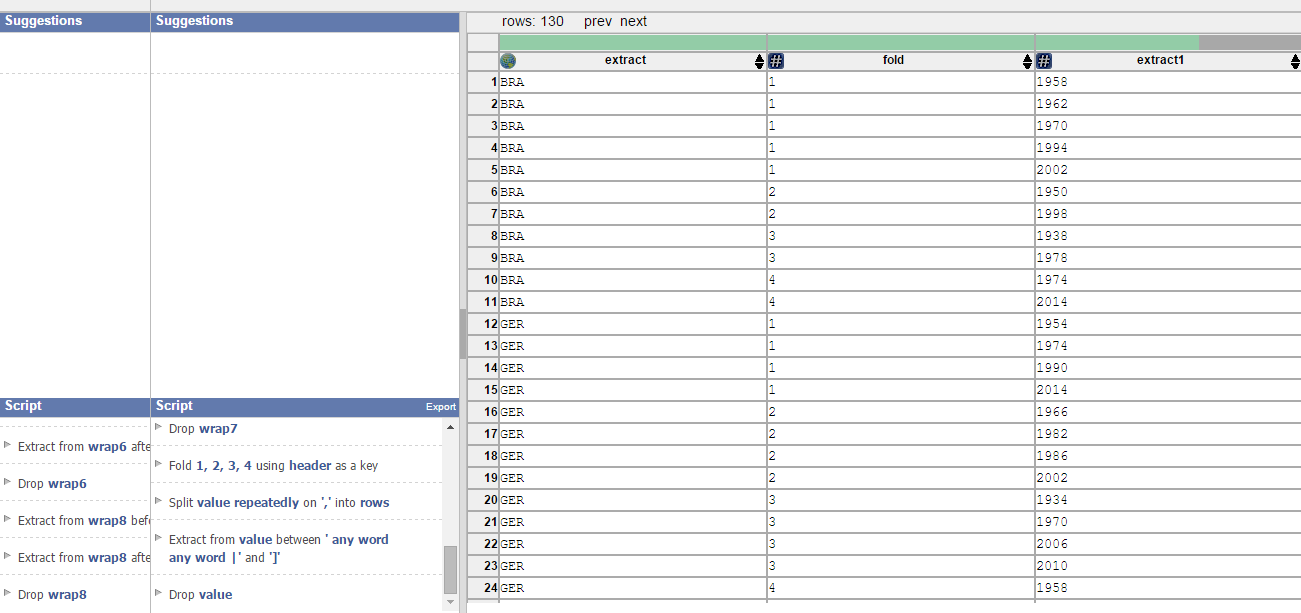
**w.add(dw.Drop(column=["value"],**

**table=0,**

**status="active",**

**drop=True))**

**w.apply\_to\_file(sys.argv[1]).print\_csv(sys.argv[2])**

****

**Data Wrangler Script and Screenshot: World Cup 2**

**from wrangler import dw**

**import sys**

**if(len(sys.argv) < 3):**

**sys.exit('Error: Please include an input and output file. Example python script.py input.csv output.csv')**

**w = dw.DataWrangler()**

**# Split data repeatedly on newline into rows**

**w.add(dw.Split(column=["data"],**

**table=0,**

**status="active",**

**drop=True,**

**result="row",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on="\n",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=0,**

**positions=None,**

**quote\_character=None))**

**# Delete row 1**

**w.add(dw.Filter(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**row=dw.Row(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**conditions=[dw.RowIndex(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**indices=[0])])))**

**# Wrap rows where data = '|-'**

**w.add(dw.Wrap(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**row=dw.Row(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**conditions=[dw.Eq(column=[],**

**table=0,**

**status="active",**

**drop=False,**

**lcol="data",**

**value="|-",**

**op\_str="=")])))**

**# Extract from wrap1 on ' any uppercase word '**

**w.add(dw.Extract(column=["wrap1"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on="[A-Z]+",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Drop wrap1**

**w.add(dw.Drop(column=["wrap1"],**

**table=0,**

**status="active",**

**drop=True))**

**# Drop wrap**

**w.add(dw.Drop(column=["wrap"],**

**table=0,**

**status="active",**

**drop=True))**

**# Set wrap2 name to 1**

**w.add(dw.SetName(column=["wrap2"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["1"],**

**header\_row=None))**

**# Set wrap3 name to 2**

**w.add(dw.SetName(column=["wrap3"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["2"],**

**header\_row=None))**

**# Set wrap4 name to 3**

**w.add(dw.SetName(column=["wrap4"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["3"],**

**header\_row=None))**

**# Set wrap5 name to 4**

**w.add(dw.SetName(column=["wrap5"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["4"],**

**header\_row=None))**

**# Drop wrap6**

**w.add(dw.Drop(column=["wrap6"],**

**table=0,**

**status="active",**

**drop=True))**

**# Drop wrap7**

**w.add(dw.Drop(column=["wrap7"],**

**table=0,**

**status="active",**

**drop=True))**

**# Fold 1, 2, 3, 4 using header as a key**

**w.add(dw.Fold(column=["\_1","\_2","\_3","\_4"],**

**table=0,**

**status="active",**

**drop=False,**

**keys=[-1]))**

**# Split value repeatedly on ',' into rows**

**w.add(dw.Split(column=["value"],**

**table=0,**

**status="active",**

**drop=True,**

**result="row",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on=",",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max="0",**

**positions=None,**

**quote\_character=None))**

**# Extract from value between ' any word any word |' and ']'**

**w.add(dw.Extract(column=["value"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

**insert\_position="right",**

**row=None,**

**on=".\*",**

**before="]",**

**after="[a-zA-Z]+[a-zA-Z]+\\|",**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Drop value**

**w.add(dw.Drop(column=["value"],**

**table=0,**

**status="active",**

**drop=True))**

**# Set fold name to place**

**w.add(dw.SetName(column=["fold"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["place"],**

**header\_row=None))**

**# Set extract1 name to year**

**w.add(dw.SetName(column=["extract1"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["year"],**

**header\_row=None))**

**# Set extract name to country**

**w.add(dw.SetName(column=["extract"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["country"],**

**header\_row=None))**

**# Unfold year, place on place**

**w.add(dw.Unfold(column=["year","place"],**

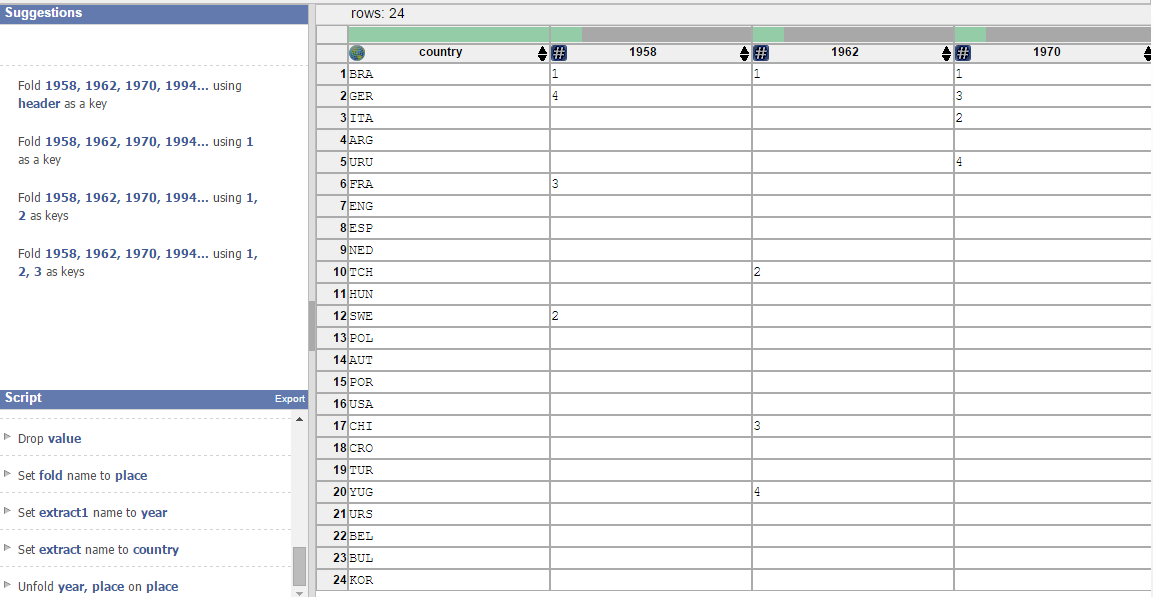
**table=0,**

**status="active",**

**drop=False,**

**measure="place"))**

**w.apply\_to\_file(sys.argv[1]).print\_csv(sys.argv[2])**

****

**UNIX Tools Command: CMSC**

**UNIX Tools Command: World Cup 1**

**Python Script: CMSC**

**Python Script: World Cup 1**

**Python Script: World Cup 2**