**Lab 4 Submission**

Name: Leah Xu UID: 112428303

**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))

# Wrap rows where data contains 'CMSC'

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

table=0,

status="active",

drop=False,

row=dw.Row(column=[],

table=0,

status="active",

drop=False,

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

table=0,

status="active",

drop=False,

lcol="data",

value="CMSC",

op\_str="contains")])))

# Drop wrap6, wrap7, wrap8, wrap9...

w.add(dw.Drop(column=["wrap6","wrap7","wrap8","wrap9","wrap10","wrap11","wrap12","wrap13","wrap14","wrap15","wrap16","wrap17","wrap18","wrap19","wrap20","wrap21","wrap22","wrap23","wrap24","wrap25","wrap26","wrap27","wrap28","wrap29","wrap30","wrap31","wrap32","wrap33","wrap34","wrap35","wrap36","wrap37","wrap38","wrap39"],

table=0,

status="active",

drop=True))

# Drop wrap73, wrap40, wrap41, wrap42...

w.add(dw.Drop(column=["wrap73","wrap40","wrap41","wrap42","wrap43","wrap44","wrap45","wrap46","wrap47","wrap48","wrap49","wrap50","wrap51","wrap52","wrap53","wrap54","wrap55","wrap56","wrap57","wrap58","wrap59","wrap60","wrap61","wrap62","wrap63","wrap64","wrap65","wrap66","wrap67","wrap68","wrap69","wrap70","wrap71","wrap72"],

table=0,

status="active",

drop=True))

# Drop wrap74, wrap75, wrap76, wrap77...

w.add(dw.Drop(column=["wrap74","wrap75","wrap76","wrap77","wrap78","wrap79","wrap80","wrap81"],

table=0,

status="active",

drop=True))

# Set wrap name to Course\_No.

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

table=0,

status="active",

drop=True,

names=["Course\_No."],

header\_row=None))

# Set wrap1 name to Section\_No.

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

table=0,

status="active",

drop=True,

names=["Section\_No."],

header\_row=None))

# Set wrap2 name to Instructor

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

table=0,

status="active",

drop=True,

names=["Instructor"],

header\_row=None))

# Extract from wrap5 on ' any number '

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

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on="\\d+",

before=None,

after=None,

ignore\_between=None,

which=1,

max=1,

positions=None))

# Extract from wrap5 on ' any word '

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

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on="[a-zA-Z]+",

before=None,

after=None,

ignore\_between=None,

which=1,

max=1,

positions=None))

# Set extract1 name to Buldg.

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

table=0,

status="active",

drop=True,

names=["Buldg."],

header\_row=None))

# Set Buldg. name to Bldg.

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

table=0,

status="active",

drop=True,

names=["Bldg."],

header\_row=None))

# Drop wrap5

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

table=0,

status="active",

drop=True))

# Extract from wrap4 after ' any word '

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

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=None,

after="[a-zA-Z]+ ",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Set extract2 name to Time

w.add(dw.SetName(column=["extract2"],

table=0,

status="active",

drop=True,

names=["Time"],

header\_row=None))

# Extract from wrap4 on ' any word '

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

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on="[a-zA-Z]+",

before=None,

after=None,

ignore\_between=None,

which=1,

max=1,

positions=None))

# Set extract3 name to Days

w.add(dw.SetName(column=["extract3"],

table=0,

status="active",

drop=True,

names=["Days"],

header\_row=None))

# Drop wrap4

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

table=0,

status="active",

drop=True))

# Extract from wrap3 between 'Waitlist: ' and ')'

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

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))

# Set extract4 name to Waitlist

w.add(dw.SetName(column=["extract4"],

table=0,

status="active",

drop=True,

names=["Waitlist"],

header\_row=None))

# Extract from wrap3 between ' Open: ' and ','

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

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))

# Set extract5 name to Open

w.add(dw.SetName(column=["extract5"],

table=0,

status="active",

drop=True,

names=["Open"],

header\_row=None))

# Extract from wrap3 between ': ' and ','

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

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))

# Set extract6 name to Seats

w.add(dw.SetName(column=["extract6"],

table=0,

status="active",

drop=True,

names=["Seats"],

header\_row=None))

# Drop wrap3

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

table=0,

status="active",

drop=True))

# Set extract name to Room\_No.

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

table=0,

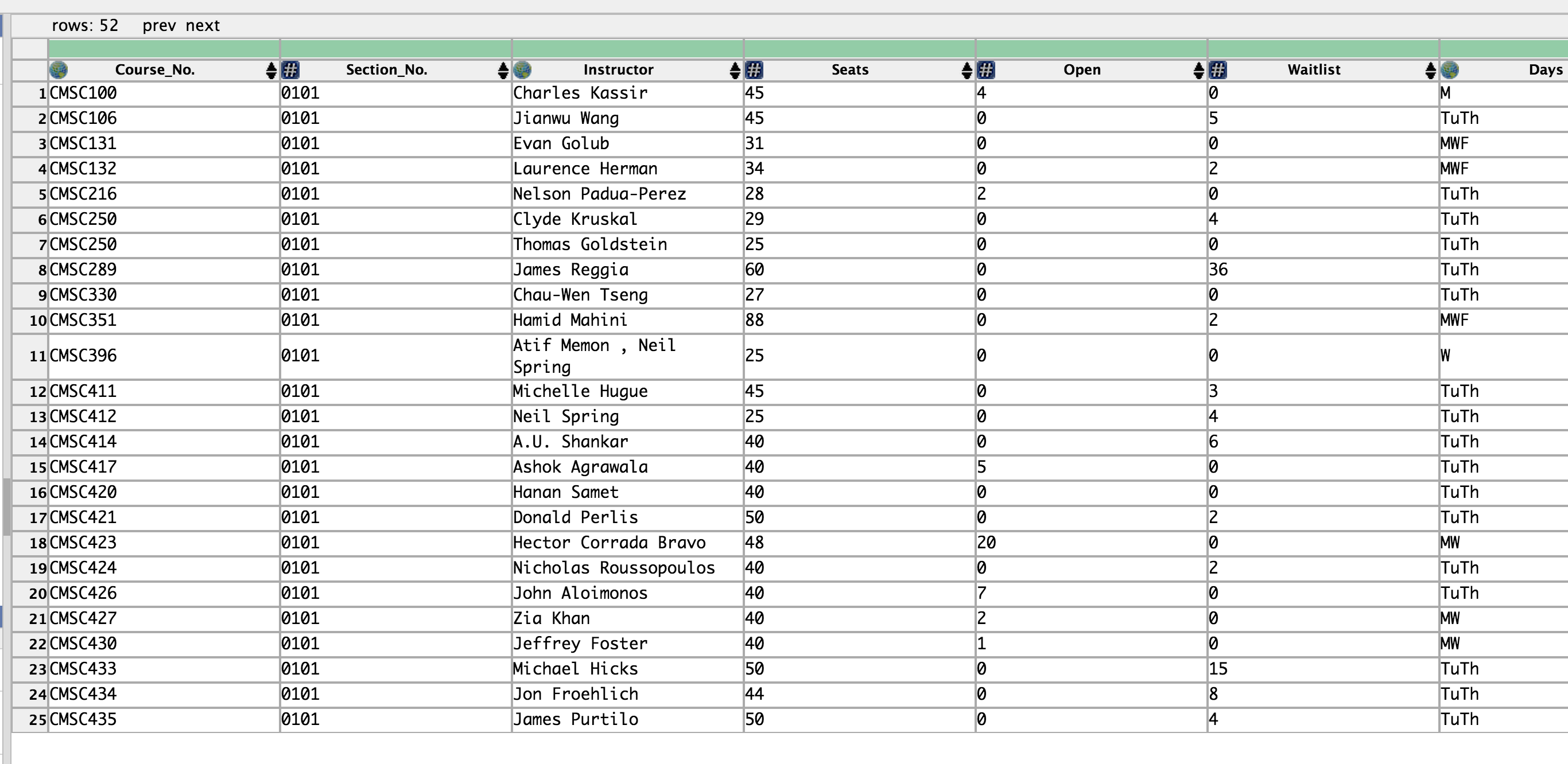
status="active",

drop=True,

names=["Room\_No."],

header\_row=None))

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))**

**# Wrap rows where data starts with '|{{fb|'**

**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="|{{fb|",**

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

**# Extract from wrap between ' any lowercase word |' and '}'**

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

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

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

**before="}",**

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

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Set extract name to Country**

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

**table=0,**

**status="active",**

**drop=True,**

**names=["Country"],**

**header\_row=None))**

**# Set Country name to Team**

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

**table=0,**

**status="active",**

**drop=True,**

**names=["Team"],**

**header\_row=None))**

**# Set wrap1 name to 1**

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

**table=0,**

**status="active",**

**drop=True,**

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

**header\_row=None))**

**# Set wrap2 name to 2**

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

**table=0,**

**status="active",**

**drop=True,**

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

**header\_row=None))**

**# Set wrap3 name to 3**

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

**table=0,**

**status="active",**

**drop=True,**

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

**header\_row=None))**

**# Set wrap4 name to 4**

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

**table=0,**

**status="active",**

**drop=True,**

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

**header\_row=None))**

**# Drop wrap**

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

**table=0,**

**status="active",**

**drop=True))**

**# Extract from 1 between '[\[' and ']'**

**w.add(dw.Extract(column=["\_1"],**

**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 1 on '2002'**

**w.add(dw.Extract(column=["\_1"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on="2002",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Extract from 1 between positions 36, 40**

**w.add(dw.Extract(column=["\_1"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[36,40]))**

**# Extract from 1 between positions 26, 30**

**w.add(dw.Extract(column=["\_1"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[26,30]))**

**# Extract from 1 between positions 16, 20**

**w.add(dw.Extract(column=["\_1"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[16,20]))**

**# Extract from 1 between positions 6, 10**

**w.add(dw.Extract(column=["\_1"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[6,10]))**

**# Set extract5 name to 1**

**w.add(dw.SetName(column=["extract5"],**

**table=0,**

**status="active",**

**drop=True,**

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

**header\_row=None))**

**# Set extract4 name to 1**

**w.add(dw.SetName(column=["extract4"],**

**table=0,**

**status="active",**

**drop=True,**

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

**header\_row=None))**

**# Set 1 name to**

**w.add(dw.SetName(column=["\_1"],**

**table=0,**

**status="active",**

**drop=True,**

**names=[""],**

**header\_row=None))**

**# Set 11 name to first1**

**w.add(dw.SetName(column=["\_11"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["first1"],**

**header\_row=None))**

**# Set 12 name to first2**

**w.add(dw.SetName(column=["\_12"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["first2"],**

**header\_row=None))**

**# Set extract3 name to first3**

**w.add(dw.SetName(column=["extract3"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["first3"],**

**header\_row=None))**

**# Set extract2 name to first4**

**w.add(dw.SetName(column=["extract2"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["first4"],**

**header\_row=None))**

**# Set extract1 name to first5**

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

**table=0,**

**status="active",**

**drop=True,**

**names=["first5"],**

**header\_row=None))**

**# Drop**

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

**table=0,**

**status="active",**

**drop=True))**

**# Drop extract**

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

**table=0,**

**status="active",**

**drop=True))**

**# Extract from 2 on '2002'**

**w.add(dw.Extract(column=["\_2"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on="2002",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Extract from 2 between positions 26, 30**

**w.add(dw.Extract(column=["\_2"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[26,30]))**

**# Extract from 2 between positions 16, 20**

**w.add(dw.Extract(column=["\_2"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[16,20]))**

**# Extract from 2 between positions 6, 10**

**w.add(dw.Extract(column=["\_2"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[6,10]))**

**# Drop 2**

**w.add(dw.Drop(column=["\_2"],**

**table=0,**

**status="active",**

**drop=True))**

**# Set extract8 name to second1**

**w.add(dw.SetName(column=["extract8"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["second1"],**

**header\_row=None))**

**# Set extract7 name to second2**

**w.add(dw.SetName(column=["extract7"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["second2"],**

**header\_row=None))**

**# Set extract6 name to second3**

**w.add(dw.SetName(column=["extract6"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["second3"],**

**header\_row=None))**

**# Set extract name to second4**

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

**table=0,**

**status="active",**

**drop=True,**

**names=["second4"],**

**header\_row=None))**

**# Extract from 3 on '2010'**

**w.add(dw.Extract(column=["\_3"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on="2010",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=None))**

**# Extract from 3 between positions 26, 30**

**w.add(dw.Extract(column=["\_3"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[26,30]))**

**# Extract from 3 between ' [\[' and ']]'**

**w.add(dw.Extract(column=["\_3"],**

**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 3 between '[\[' and ']'**

**w.add(dw.Extract(column=["\_3"],**

**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))**

**# Drop 3**

**w.add(dw.Drop(column=["\_3"],**

**table=0,**

**status="active",**

**drop=True))**

**# Set extract11 name to third1**

**w.add(dw.SetName(column=["extract11"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["third1"],**

**header\_row=None))**

**# Set extract10 name to third2**

**w.add(dw.SetName(column=["extract10"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["third2"],**

**header\_row=None))**

**# Set extract9 name to third3**

**w.add(dw.SetName(column=["extract9"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["third3"],**

**header\_row=None))**

**# Set extract name to third4**

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

**table=0,**

**status="active",**

**drop=True,**

**names=["third4"],**

**header\_row=None))**

**# Extract from 4 between positions 26, 30**

**w.add(dw.Extract(column=["\_4"],**

**table=0,**

**status="active",**

**drop=False,**

**result="column",**

**update=False,**

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

**row=None,**

**on=None,**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=1,**

**positions=[26,30]))**

**# Extract from 4 between ' [\[' and ']'**

**w.add(dw.Extract(column=["\_4"],**

**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 4 between '[\[' and ']'**

**w.add(dw.Extract(column=["\_4"],**

**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))**

**# Drop 4**

**w.add(dw.Drop(column=["\_4"],**

**table=0,**

**status="active",**

**drop=True))**

**# Set extract13 name to fourth1**

**w.add(dw.SetName(column=["extract13"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["fourth1"],**

**header\_row=None))**

**# Set extract12 name to fourth2**

**w.add(dw.SetName(column=["extract12"],**

**table=0,**

**status="active",**

**drop=True,**

**names=["fourth2"],**

**header\_row=None))**

**# Set extract name to fourth3**

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

**table=0,**

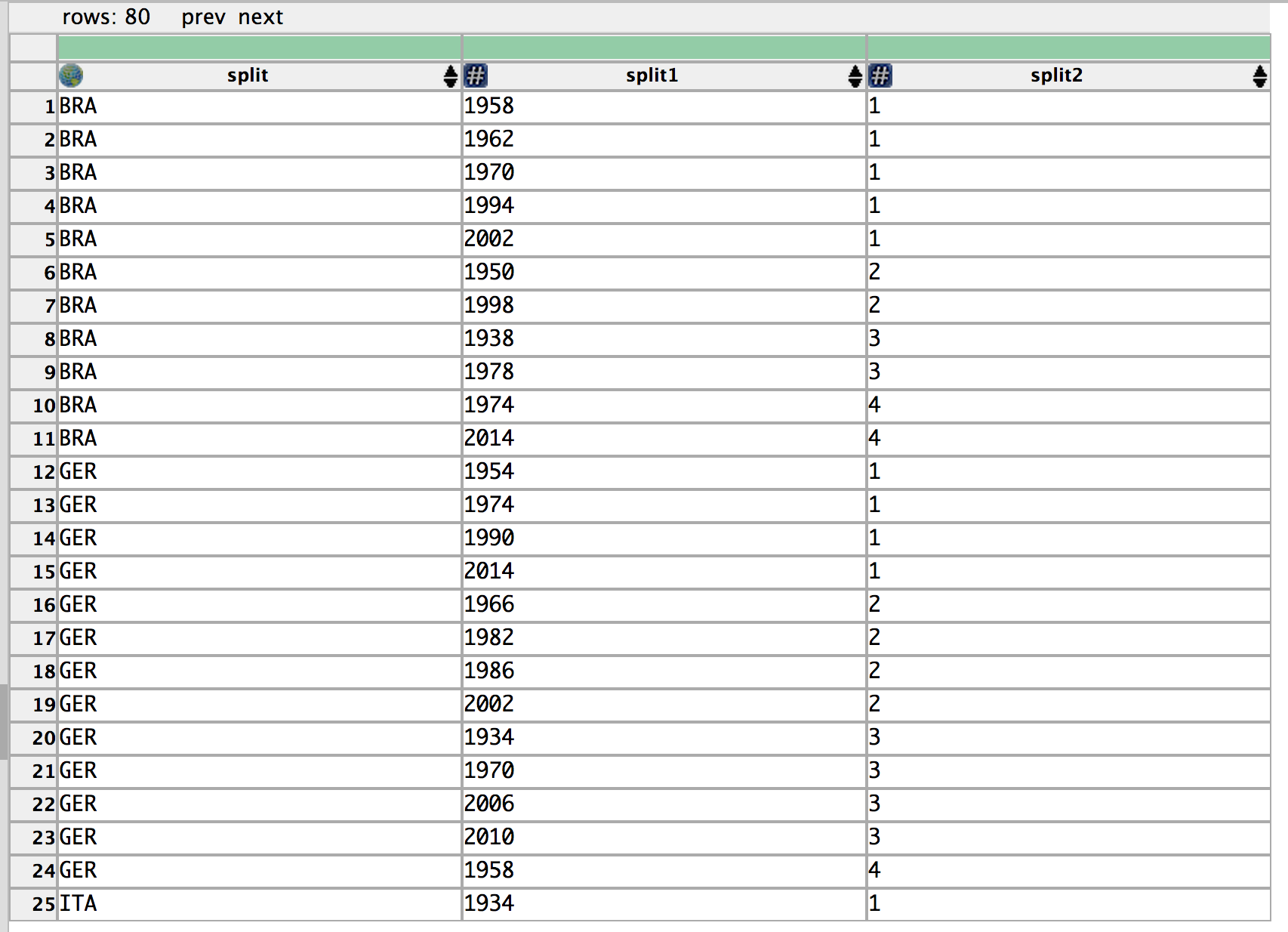
**status="active",**

**drop=True,**

**names=["fourth3"],**

**header\_row=None))**

**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))**

**# Split data repeatedly on ','**

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

**table=0,**

**status="active",**

**drop=True,**

**result="column",**

**update=False,**

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

**row=None,**

**on=",",**

**before=None,**

**after=None,**

**ignore\_between=None,**

**which=1,**

**max=0,**

**positions=None,**

**quote\_character=None))**

**# Unfold split1 on split2**

**w.add(dw.Unfold(column=["split1"],**

**table=0,**

**status="active",**

**drop=False,**

**measure="split2"))**

**# Set split name to team**

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

**table=0,**

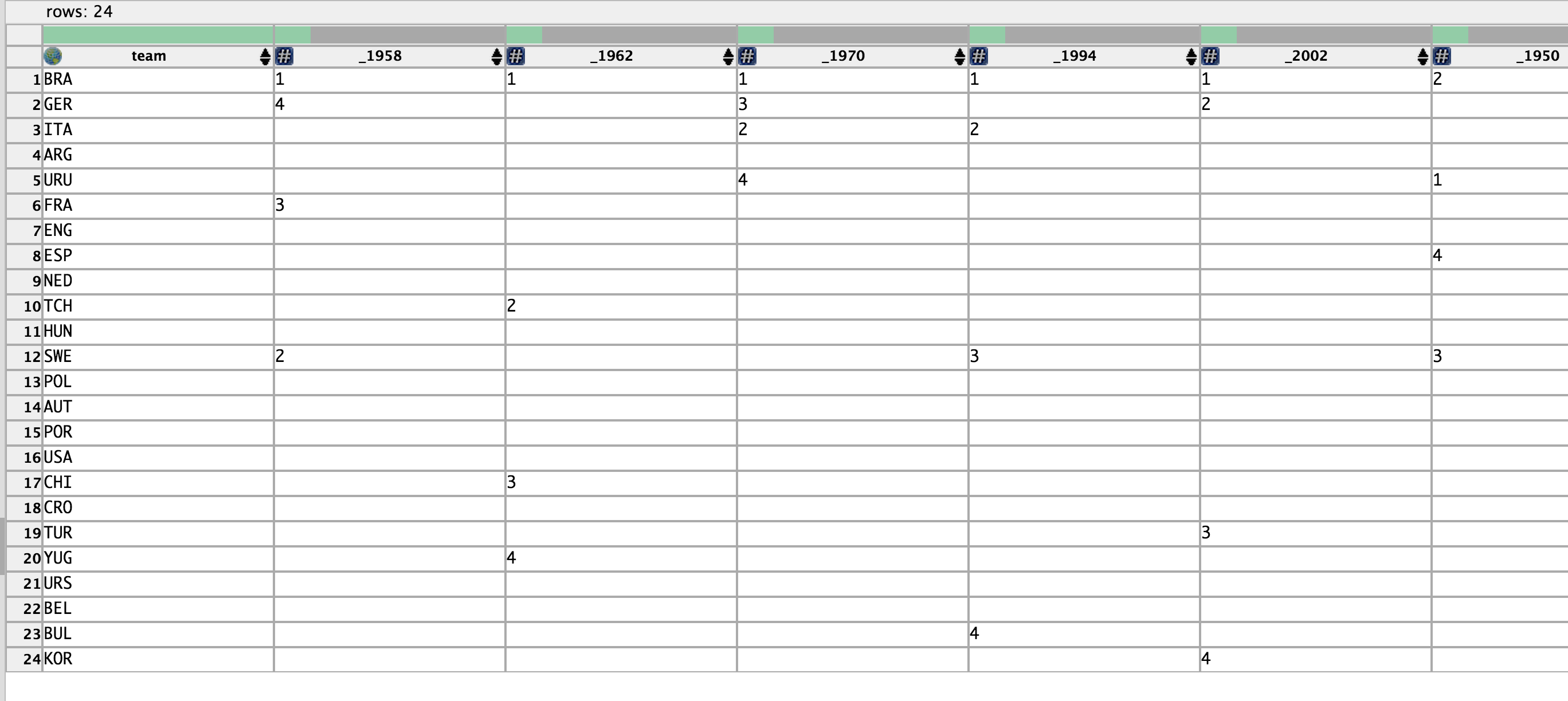
**status="active",**

**drop=True,**

**names=["team"],**

**header\_row=None))**

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

****

**UNIX Tools Command: CMSC**

**cat cmsc.txt | awk -F',' 'BEGIN {printf "Course No., Section No., Instructor, Seats, Open, Waitlist, Days, Time, Bldg., Room No."}**

**/^CMSC/ {print combined; combined = $0}**

**!/^CMSC/ {combined = combined", "$0;}**

**END {print combined}' | sed 's/Seats (Total: //g; s/Open: //g;**

**s/Waitlist: //g; s/)//g; s/ /, /g; s/, $//g;**

**s/MWF/MWF,/g; s/MW /MW, /g; s/TuTh/TuTh,/g; s/M /M, /g;**

**s/Tu /Tu, /g; s/W /W, /g; s/Th /Th, /g; s/F /F, /g; s/ ,/,/g' > partII\_cmsc.csv**

**UNIX Tools Command: World Cup 1**

**cat worldcup.txt | sed '**

**s/|align=center|{{sort dash}}/none/g**

**s/||{{sort dash}}//g;**

**s/|-//g;**

**s/|style="background:#fff68f"//g;**

**s/!.\*finishes//g;**

**s/|style=white-space:nowrap//g;**

**s/<.\*sup>//g;**

**s/\[\[#1|\\*\]\]//g;**

**s/GER[[#2|^]]//g;**

**s/|.$//g;**

**s/|[[:digit:]]|//g;**

**s/[[:digit:]]\* FIFA World Cup|//g;**

**s/\[\[#2\|\^\]\]//g;**

**s/|11|//g;**

**s/|13||12|//g;**

**s/\|{{fb\|//g;**

**s/|[0-9]//g;**

**s/|//g; s/}//g;**

**s/\[//g;**

**s/\]//g;**

**s/(//g;**

**s/)//g;**

**s/, /,/g;**

**s/ //g' | sed '/^$/d' | awk '**

**/^[A-Z]/ {team = $1}**

**!/^[[A-Z]/ {print team; split($0,a,",");**

**for (i=1; i<=5; i++) {**

**if (a[i] != 0)**

**printf("%d, %d\n", a[i], (NR-1)%5)**

**}**

**}' | awk '/^[A-Z]/ {team = $1} /^[1-9]/ {print team", "$0}' > partII\_worldcup1.csv**

**Python Script: CMSC**

**import re**

**with open("cmsc.txt", "r") as sources:**

**lines = sources.readlines()**

**with open("partIII\_cmsc.csv", "w") as sources:**

**sources.write("Course No., Section No., Instructor, Seats, Open, Waitlist, Days, Time, Bldg., Room No.")**

**course = ''**

**info = ''**

**for line in lines:**

**if line.find("CMSC") == -1:**

**line = re.sub(r'Seats \(Total: ', '', line)**

**line = re.sub(r'Open: ', '', line)**

**line = re.sub(r'Waitlist: ', '', line)**

**line = re.sub(r'\)', '', line)**

**line = re.sub(r'MWF', 'MWF,', line)**

**line = re.sub(r'MW ', 'MW,', line)**

**line = re.sub(r'TuTh', 'TuTh,', line)**

**line = re.sub(r'M ', 'M, ', line)**

**line = re.sub(r'Tu ', 'Tu, ', line)**

**line = re.sub(r'W ', 'W, ', line)**

**line = re.sub(r'Th ', 'Th, ', line)**

**line = re.sub(r'F ', 'F, ', line)**

**line = re.sub(r' ', ', ', line)**

**line = line.rstrip()**

**info = info + ', ' + line**

**else:**

**sources.write(course.rstrip() + info[:-2] + '\n')**

**course = line**

**info = ''**

**Course No., Section No., Instructor, Seats, Open, Waitlist, Days, Time, Bldg., Room No.**

**CMSC100, 0101, Charles Kassir, 45, 4, 0, M, 4:00pm - 4:50pm, CSI, 2117**

**CMSC106, 0101, Jianwu Wang, 45, 0, 5, TuTh, 9:30am - 10:45am, CSI, 2117**

**CMSC131, 0101, Evan Golub, 31, 0, 0, MWF, 2:00pm - 2:50pm, CSI, 1115, 0102, Evan Golub, 31, 0, 0, MWF, 2:00pm - 2:50pm, CSI, 1115, 0103, Evan Golub, 31, 1, 0, MWF, 2:00pm - 2:50pm, CSI, 1115, 0104, Evan Golub, 31, 1, 0, MWF, 2:00pm - 2:50pm, CSI, 1115, 0201, Evan Golub, 31, 1, 0, MWF, 3:00pm - 3:50pm, CSI, 1115, 0202, Evan Golub, 31, 1, 0, MWF, 3:00pm - 3:50pm, CSI, 1115, 0203, Evan Golub, 31, 0, 0, MWF, 3:00pm - 3:50pm, CSI, 1115, 0204, Evan Golub, 31, 0, 1, MWF, 3:00pm - 3:50pm, CSI, 1115, 0301, Thomas Reinhardt, 31, 0, 0, MWF, 11:00am - 11:50am, CSI, 1115, 0302, Thomas Reinhardt, 31, 0, 0, MWF, 11:00am - 11:50am, CSI, 1115, 0303, Thomas Reinhardt, 31, 0, 0, MWF, 11:00am - 11:50am, CSI, 1115, 0304, Thomas Reinhardt, 31, 0, 0, MWF, 11:00am - 11:50am, CSI, 1115, 0401, Thomas Reinhardt, 31, 0, 0, MWF, 12:00pm - 12:50pm, CSI, 1115, 0402, Thomas Reinhardt, 31, 7, 0, MWF, 12:00pm - 12:50pm, CSI, 1115, 0403, Thomas Reinhardt, 31, 0, 0, MWF, 12:00pm - 12:50pm, CSI, 1115, 0404, Thomas Reinhardt, 31, 7, 0, MWF, 12:00pm - 12:50pm, CSI, 1115**

**CMSC132, 0101, Laurence Herman, 34, 0, 2, MWF, 10:00am - 10:50am, CSI, 1115, 0102, Laurence Herman, 34, 0, 0, MWF, 10:00am - 10:50am, CSI, 1115, 0103, Laurence Herman, 34, 0, 0, MWF, 10:00am - 10:50am, CSI, 1115, 0104, Laurence Herman, 34, 0, 2, MWF, 10:00am - 10:50am, CSI, 1115, 0201, Laurence Herman, 34, 6, 0, MWF, 1:00pm - 1:50pm, CSI, 1115, 0202, Laurence Herman, 34, 1, 0, MWF, 1:00pm - 1:50pm, CSI, 1115, 0203, Laurence Herman, 34, 0, 0, MWF, 1:00pm - 1:50pm, CSI, 1115, 0204, Laurence Herman, 34, 0, 0, MWF, 1:00pm - 1:50pm, CSI, 1115, 0301, Laurence Herman, 30, 3, 0, MWF, 2:00pm - 2:50pm, CSI, 2117, 0302, Laurence Herman, 29, 0, 1, MWF, 2:00pm - 2:50pm, CSI, 2117, 0303, Laurence Herman, 29, 8, 0, MWF, 2:00pm - 2:50pm, CSI, 2117**

**CMSC216, 0101, Nelson Padua-Perez, 28, 2, 0, TuTh, 9:30am - 10:45am, CSI, 1115, 0102, Nelson Padua-Perez, 28, 0, 0, TuTh, 9:30am - 10:45am, CSI, 1115, 0103, Nelson Padua-Perez, 28, 0, 1, TuTh, 9:30am - 10:45am, CSI, 1115, 0104, Nelson Padua-Perez, 28, 0, 0, TuTh, 9:30am - 10:45am, CSI, 1115, 0201, Nelson Padua-Perez, 28, 0, 0, TuTh, 11:00am - 12:15pm, CSI, 1115, 0202, Nelson Padua-Perez, 28, 0, 1, TuTh, 11:00am - 12:15pm, CSI, 1115, 0203, Nelson Padua-Perez, 28, 0, 1, TuTh, 11:00am - 12:15pm, CSI, 1115, 0204, Nelson Padua-Perez, 28, 0, 0, TuTh, 11:00am - 12:15pm, CSI, 1115, 0301, Nelson Padua-Perez, 28, 0, 0, TuTh, 2:00pm - 3:15pm, CSI, 1115, 0302, Nelson Padua-Perez, 28, 2, 0, TuTh, 2:00pm - 3:15pm, CSI, 1115, 0303, Nelson Padua-Perez, 28, 0, 0, TuTh, 2:00pm - 3:15pm, CSI, 1115, 0304, Nelson Padua-Perez, 28, 0, 0, TuTh, 2:00pm - 3:15pm, CSI, 1115**

**CMSC250, 0101, Clyde Kruskal, 29, 0, 4, TuTh, 2:00pm - 3:15pm, CSI, 2117, 0102, Clyde Kruskal, 29, 0, 5, TuTh, 2:00pm - 3:15pm, CSI, 2117, 0103, Clyde Kruskal, 29, 0, 3, TuTh, 2:00pm - 3:15pm, CSI, 2117, 0201, Clyde Kruskal, 29, 0, 5, TuTh, 3:30pm - 4:45pm, CSI, 1115, 0202, Clyde Kruskal, 29, 0, 3, TuTh, 3:30pm - 4:45pm, CSI, 1115, 0203, Clyde Kruskal, 29, 0, 3, TuTh, 3:30pm - 4:45pm, CSI, 1115, 0204, Clyde Kruskal, 29, 14, 0, TuTh, 3:30pm - 4:45pm, CSI, 1115, 0301, Thomas Reinhardt, 29, 0, 1, TuTh, 11:00am - 12:15pm, CSI, 3117, 0302, Thomas Reinhardt, 29, 1, 0, TuTh, 11:00am - 12:15pm, CSI, 3117, 0303, Thomas Reinhardt, 29, 0, 5, TuTh, 11:00am - 12:15pm, CSI, 3117**

**CMSC250, 0101, Thomas Goldstein, 25, 0, 0, TuTh, 12:30pm - 1:45pm, CSI, 3118**

**CMSC289, 0101, James Reggia, 60, 0, 36, TuTh, 11:00am - 12:15pm, CSI, 2117**

**CMSC330, 0101, Chau-Wen Tseng, 27, 0, 0, TuTh, 3:30pm - 4:45pm, CSI, 3117, 0102, Chau-Wen Tseng, 27, 0, 0, TuTh, 3:30pm - 4:45pm, CSI, 3117, 0103, Chau-Wen Tseng, 27, 1, 0, TuTh, 3:30pm - 4:45pm, CSI, 3117, 0201, Chau-Wen Tseng, 27, 0, 0, TuTh, 12:30pm - 1:45pm, CSI, 3117, 0202, Chau-Wen Tseng, 27, 0, 1, TuTh, 12:30pm - 1:45pm, CSI, 3117, 0203, Chau-Wen Tseng, 27, 13, 0, TuTh, 12:30pm - 1:45pm, CSI, 3117, 0301, Chau-Wen Tseng, 27, 0, 0, TuTh, 2:00pm - 3:15pm, CSI, 3117, 0302, Chau-Wen Tseng, 27, 0, 0, TuTh, 2:00pm - 3:15pm, CSI, 3117, 0303, Chau-Wen Tseng, 27, 1, 0, TuTh, 2:00pm - 3:15pm, CSI, 3117**

**CMSC351, 0101, Hamid Mahini, 88, 0, 2, MWF, 10:00am - 10:50am, CSI, 3117, 0201, Hamid Mahini, 88, 0, 4, MWF, 11:00am - 11:50am, CSI, 3117, 0301, Hamid Mahini, 88, 0, 0, MWF, 3:00pm - 3:50pm, CSI, 2117**

**CMSC396, 0101, Atif Memon , Neil Spring, 25, 0, 0, W, 1:00pm - 1:50pm, AVW,, 4172**

**CMSC411, 0101, Michelle Hugue, 45, 0, 3, TuTh, 3:30pm - 4:45pm, CSI, 1122, 0201, Michelle Hugue, 45, 0, 3, TuTh, 2:00pm - 3:15pm, CSI, 1122**

**CMSC412, 0101, Neil Spring, 25, 0, 4, TuTh, 11:00am - 12:15pm, CSI, 1122, 0102, Neil Spring, 25, 0, 3, TuTh, 11:00am - 12:15pm, CSI, 1122**

**CMSC414, 0101, A.U. Shankar, 40, 0, 6, TuTh, 12:30pm - 1:45pm, CSI, 1122, 0201, Elaine Shi, 50, 1, 0, MW,3:30pm - 4:45pm, CSI, 1122**

**CMSC417, 0101, Ashok Agrawala, 40, 5, 0, TuTh, 11:00am - 12:15pm, CSI, 3120, 0201, Instructor: TBA, 40, 27, 0, TuTh, 3:30pm - 4:45pm, CSI, 2107**

**CMSC420, 0101, Hanan Samet, 40, 0, 0, TuTh, 12:30pm - 1:45pm, CSI, 2120, 0201, Michelle Hugue, 40, 0, 10, MW,3:30pm - 4:45pm, CSI, 3120**

**CMSC421, 0101, Donald Perlis, 50, 0, 2, TuTh, 2:00pm - 3:15pm, CSI, 1121, 0201, Donald Perlis, 46, 0, 3, TuTh, 9:30am - 10:45am, CSI, 3117**

**CMSC423, 0101, Hector Corrada Bravo, 48, 20, 0, MW,3:30pm - 4:45pm, CSI, 1121**

**CMSC424, 0101, Nicholas Roussopoulos, 40, 0, 2, TuTh, 2:00pm - 3:15pm, CSI, 3120**

**CMSC426, 0101, John Aloimonos, 40, 7, 0, TuTh, 11:00am - 12:15pm, CSI, 2120**

**CMSC427, 0101, Zia Khan, 40, 2, 0, MW,2:00pm - 3:15pm, CSI, 1121**

**CMSC430, 0101, Jeffrey Foster, 40, 1, 0, MW,3:30pm - 4:45pm, CSI, 2107**

**CMSC433, 0101, Michael Hicks, 50, 0, 15, TuTh, 3:30pm - 4:45pm, CSI, 1121**

**CMSC434, 0101, Jon Froehlich, 44, 0, 8, TuTh, 9:30am - 10:45am, CSI, 1122, 0201, Vibha Sazawal, 44, 0, 6, TuTh, 11:00am - 12:15pm, CSI, 1121**

**CMSC435, 0101, James Purtilo, 50, 0, 4, TuTh, 12:30pm - 1:45pm, CSI, 1121**

**CMSC436, 0101, Adam Porter, 50, 0, 14, TuTh, 9:30am - 10:45am, CSI, 1121, 0201, Atif Memon, 60, 0, 11, TuTh, 12:30pm - 1:45pm, CSI, 2117**

**CMSC451, 0101, Aravind Srinivasan, 40, 0, 1, TuTh, 12:30pm - 1:45pm, CSI, 3120**

**CMSC452, 0101, William Gasarch, 32, 11, 0, TuTh, 2:00pm - 3:15pm, CSI, 3118**

**CMSC456, 0101, Jonathan Katz, 40, 0, 3, MWF, 12:00pm - 12:50pm, CSI, 1121**

**CMSC460, 0101, Harland Glaz, 33, 0, 2, TuTh, 12:30pm - 1:45pm, MTH, 0304, 0201, Changhui Tan, 25, 2, 0, TuTh, 2:00pm - 3:15pm, MTH, 0409**

**CMSC466, 0101, Maria Cameron, 25, 11, 0, MWF, 11:00am - 11:50am, MTH, 1311**

**CMSC474, 0101, Dana Nau, 40, 9, 0, TuTh, 3:30pm - 4:45pm, CSI, 2120**

**CMSC498, 0101, Uzi Vishkin, 10, 7, 0, MW,11:00am - 12:15pm, ITV, 1100**

**CMSC498, 0101, Amol Deshpande, 40, 8, 0, TuTh, 11:00am - 12:15pm, CSI, 2107**

**CMSC498, 0101, Marshini Chetty, 10, 0, 4, Th, 6:00pm - 8:45pm, HBK, 0123**

**CMSC631, 0101, David Van Horn, 20, 3, 0, TuTh, 3:30pm - 4:45pm, CSI, 3118**

**CMSC660, 0101, Ramani Duraiswami, 25, 0, 3, TuTh, 3:30pm - 4:45pm, CSI, 3120**

**CMSC663, 0101, Howard Elman, 20, 15, 0, TuTh, 5:00pm - 6:15pm, CSI, 4122**

**CMSC666, 0101, Tobias von Petersdorff, 25, 8, 0, TuTh, 2:00pm - 3:15pm, CSI, 4122**

**CMSC701, 0101, Mihai Pop, 40, 9, 0, TuTh, 9:30am - 10:45am, CSI, 3120**

**CMSC712, 0101, A.U. Shankar, 20, 0, 0, TuTh, 2:00pm - 3:30pm, AVW,, 3258**

**CMSC723, 0101, Hal Daume, 40, 0, 10, TuTh, 2:00pm - 3:15pm, CSI, 2120**

**CMSC733, 0101, David Jacobs, 32, 0, 17, TuTh, 11:00am - 12:15pm, CSI, 3118**

**CMSC737, 0101, Atif Memon, 20, 0, 0, TuTh, 9:30am - 10:45am, CSI, 2118**

**CMSC754, 0101, David Mount, 32, 0, 15, TuTh, 12:30pm - 1:45pm, CSI, 2118**

**CMSC798, 0101, Jeffrey Foster, 40, 10, 0, F, 11:00am - 12:50pm, CSI, 2117**

**CMSC818, 0101, Peter Keleher, 20, 1, 0, TuTh, 11:00am - 12:15pm, AVW,, 3258**

**CMSC818, 0101, Instructor: TBA, 145, 145, 0, MW,4:00pm - 5:00pm, CSI, 1115**

**CMSC818, 0101, Tudor Dumitras, 15, 13, 0, MW,12:30pm - 1:45pm, JMP, 1202**

**CMSC828, 0101, Larry Davis, 25, 2, 0, TuTh, 2:00pm - 3:15pm, CSI, 2118**

**CMSC828, 0101, Venkatramanan Subrahmanian, 25, 12, 0, TuTh, 9:30am - 10:45am, CSI, 3118**

**Python Script: World Cup 1**

**import re**

**with open("worldcup.txt", "r") as sources:**

**lines = sources.readlines()**

**with open("partIII\_worldcup1.csv", "w") as sources:**

**sources.write("Team,Year,Place\n")**

**place = 0**

**country = ''**

**for line in lines:**

**line = re.sub(r'align=center\|{{sort dash}}', 'none', line)**

**line = re.sub(r'\|\|{{sort dash}}', '', line)**

**line = re.sub(r'\|-', '', line)**

**line = re.sub(r'!.\*finishes', '', line)**

**line = re.sub(r'\|style="background:#fff68f"', '', line)**

**line = re.sub(r'\|style=white-space:nowrap', '', line)**

**line = re.sub(r'<.\*sup>', '', line)**

**line = re.sub(r'\[\[#1\|\\*\]\]', '', line)**

**line = re.sub(r'GER\[\[#2\|\^\]\]', '', line)**

**line = re.sub(r'\|.$', '', line)**

**line = re.sub(r'\d{4} FIFA World Cup\|', '', line)**

**line = re.sub(r'\|{{fb\|', '', line)**

**line = re.sub(r'}', '', line)**

**line = re.sub(r'\|\d(\|\|)\*(\d\|)\*', '', line)**

**line = re.sub(r'\|', '', line)**

**line = re.sub(r'\[', '', line)**

**line = re.sub(r'\]', '', line)**

**line = re.sub(r'\(', '', line)**

**line = re.sub(r'\)', '', line)**

**line = re.sub(r' ', '', line)**

**line = re.sub(r'#2\^', '', line)**

**line = line.rstrip()**

**m = re.match(r'[A-Z]{3}', line)**

**if m:**

**country = line**

**years = line.split(',')**

**if ((len(line) > 1) & (line != country)):**

**for year in years:**

**if (year != "none"):**

**sources.write(country + ', ' + year + ', ' + str((place % 4)+1) + '\n')**

**place = place + 1**

**BRA, 1958, 1**

**BRA, 1962, 1**

**BRA, 1970, 1**

**BRA, 1994, 1**

**BRA, 2002, 1**

**BRA, 1950, 2**

**BRA, 1998, 2**

**BRA, 1938, 3**

**BRA, 1978, 3**

**BRA, 1974, 4**

**BRA, 2014, 4**

**GER, 1954, 1**

**GER, 1974, 1**

**GER, 1990, 1**

**GER, 2014, 1**

**GER, 1966, 2**

**GER, 1982, 2**

**GER, 1986, 2**

**GER, 2002, 2**

**GER, 1934, 3**

**GER, 1970, 3**

**GER, 2006, 3**

**GER, 2010, 3**

**GER, 1958, 4**

**ITA, 1934, 1**

**ITA, 1938, 1**

**ITA, 1982, 1**

**ITA, 2006, 1**

**ITA, 1970, 2**

**ITA, 1994, 2**

**ITA, 1990, 3**

**ITA, 1978, 4**

**ARG, 1978, 1**

**ARG, 1986, 1**

**ARG, 1930, 2**

**ARG, 1990, 2**

**ARG, 2014, 2**

**URU, 1930, 1**

**URU, 1950, 1**

**URU, 1954, 4**

**URU, 1970, 4**

**URU, 2010, 4**

**FRA, 1998, 1**

**FRA, 2006, 2**

**FRA, 1958, 3**

**FRA, 1986, 3**

**FRA, 1982, 4**

**ENG, 1966, 1**

**ENG, 1990, 4**

**ESP, 2010, 1**

**ESP, 1950, 4**

**NED, 1974, 2**

**NED, 1978, 2**

**NED, 2010, 2**

**NED, 2014, 3**

**NED, 1998, 4**

**TCH, 1934, 2**

**TCH, 1962, 2**

**HUN, 1938, 2**

**HUN, 1954, 2**

**SWE, 1958, 2**

**SWE, 1950, 3**

**SWE, 1994, 3**

**SWE, 1938, 4**

**POL, 1974, 3**

**POL, 1982, 3**

**AUT, 1954, 3**

**AUT, 1934, 4**

**POR, 1966, 3**

**POR, 2006, 4**

**USA, 1930, 3**

**CHI, 1962, 3**

**CRO, 1998, 3**

**TUR, 2002, 3**

**YUG, 1930, 4**

**YUG, 1962, 4**

**URS, 1966, 4**

**BEL, 1986, 4**

**BUL, 1994, 4**

**KOR, 2002, 4**

**Python Script: World Cup 2**

**import pandas as pd**

**import numpy as np**

**worldcup = pd.read\_csv('partIII\_worldcup1.csv')**

**print worldcup.pivot(index='Teams', columns='Year', values='Place')**

**Year 1930 1934 1938 1950 1954 1958 1962 1966 1970 1974 1978 1982 \**

**Team**

**ARG 2 NaN NaN NaN NaN NaN NaN NaN NaN NaN 1 NaN**

**AUT NaN 4 NaN NaN 3 NaN NaN NaN NaN NaN NaN NaN**

**BEL NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN**

**BRA NaN NaN 3 2 NaN 1 1 NaN 1 4 3 NaN**

**BUL NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN**

**CHI NaN NaN NaN NaN NaN NaN 3 NaN NaN NaN NaN NaN**

**CRO NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN**

**ENG NaN NaN NaN NaN NaN NaN NaN 1 NaN NaN NaN NaN**

**ESP NaN NaN NaN 4 NaN NaN NaN NaN NaN NaN NaN NaN**

**FRA NaN NaN NaN NaN NaN 3 NaN NaN NaN NaN NaN 4**

**GER NaN 3 NaN NaN 1 4 NaN 2 3 1 NaN 2**

**HUN NaN NaN 2 NaN 2 NaN NaN NaN NaN NaN NaN NaN**

**ITA NaN 1 1 NaN NaN NaN NaN NaN 2 NaN 4 1**

**KOR NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN**

**NED NaN NaN NaN NaN NaN NaN NaN NaN NaN 2 2 NaN**

**POL NaN NaN NaN NaN NaN NaN NaN NaN NaN 3 NaN 3**

**POR NaN NaN NaN NaN NaN NaN NaN 3 NaN NaN NaN NaN**

**SWE NaN NaN 4 3 NaN 2 NaN NaN NaN NaN NaN NaN**

**TCH NaN 2 NaN NaN NaN NaN 2 NaN NaN NaN NaN NaN**

**TUR NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN**

**URS NaN NaN NaN NaN NaN NaN NaN 4 NaN NaN NaN NaN**

**URU 1 NaN NaN 1 4 NaN NaN NaN 4 NaN NaN NaN**

**USA 3 NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN**

**YUG 4 NaN NaN NaN NaN NaN 4 NaN NaN NaN NaN NaN**

**Year 1986 1990 1994 1998 2002 2006 2010 2014**

**Team**

**ARG 1 2 NaN NaN NaN NaN NaN 2**

**AUT NaN NaN NaN NaN NaN NaN NaN NaN**

**BEL 4 NaN NaN NaN NaN NaN NaN NaN**

**BRA NaN NaN 1 2 1 NaN NaN 4**

**BUL NaN NaN 4 NaN NaN NaN NaN NaN**

**CHI NaN NaN NaN NaN NaN NaN NaN NaN**

**CRO NaN NaN NaN 3 NaN NaN NaN NaN**

**ENG NaN 4 NaN NaN NaN NaN NaN NaN**

**ESP NaN NaN NaN NaN NaN NaN 1 NaN**

**FRA 3 NaN NaN 1 NaN 2 NaN NaN**

**GER 2 1 NaN NaN 2 3 3 1**

**HUN NaN NaN NaN NaN NaN NaN NaN NaN**

**ITA NaN 3 2 NaN NaN 1 NaN NaN**

**KOR NaN NaN NaN NaN 4 NaN NaN NaN**

**NED NaN NaN NaN 4 NaN NaN 2 3**

**POL NaN NaN NaN NaN NaN NaN NaN NaN**

**POR NaN NaN NaN NaN NaN 4 NaN NaN**

**SWE NaN NaN 3 NaN NaN NaN NaN NaN**

**TCH NaN NaN NaN NaN NaN NaN NaN NaN**

**TUR NaN NaN NaN NaN 3 NaN NaN NaN**

**URS NaN NaN NaN NaN NaN NaN NaN NaN**

**URU NaN NaN NaN NaN NaN NaN 4 NaN**

**USA NaN NaN NaN NaN NaN NaN NaN NaN**

**YUG NaN NaN NaN NaN NaN NaN NaN NaN**

**[24 rows x 20 columns]**