DSC640-T301
Chris Kellogg
Weeks 7 & 8 Exercise

Audience
My audience is the Multi-State Lottery Association (MUSL) and the New York State Gaming Commission (NYGC). These are the governing bodies that oversee the Mega Millions (MUSL), Pick 10 (NYGC), and Powerball (MUSL) lotteries. They are therefore well-versed in the operation of lotteries and the odds involved.

Purpose
I'd like to encourage the MUSL to change its Mega Millions game from its current format of choosing 5 balls out of 75 and 1 "bonus ball" out of 15 (hereafter described as 5/75+1/15) to a new format of 5/55+1/84. I'd also like to encourage the MUSL to changes its Powerball game from its current format of 5/70+1/25 to a new format of 5/55+1/87. Finally, I'd like to encourage the NYGC to change its Pick 10 game from its current format that requires a winner to choose 10 out of 80 balls and match 10 of the 20 balls chosen by the lottery (hereafter described as 10/20/80); I'd like them to use a new format of 10/15/55.

I'm advocating for these changes on behalf of the number 55. The number 55 used to be woven deeply into the fabric of all Americans, with Interstate highway speed limits set at 55 and the standard retirement age set at 55. In my lifetime, the number 55 has been relegated to the back burner; even the pro sports Halls of Fame feature very few who wore 55 on their jerseys. I'd like to see the number 55 featured by the lotteries as a way to restore some of its former glory and respect.

Medium
I'm drafting a petition to be sent to the MUSL and the NYGC, marshalling the support of other like-minded supporters of the number 55 who would like to see it occupy a place of prominence in the minds of the US public.

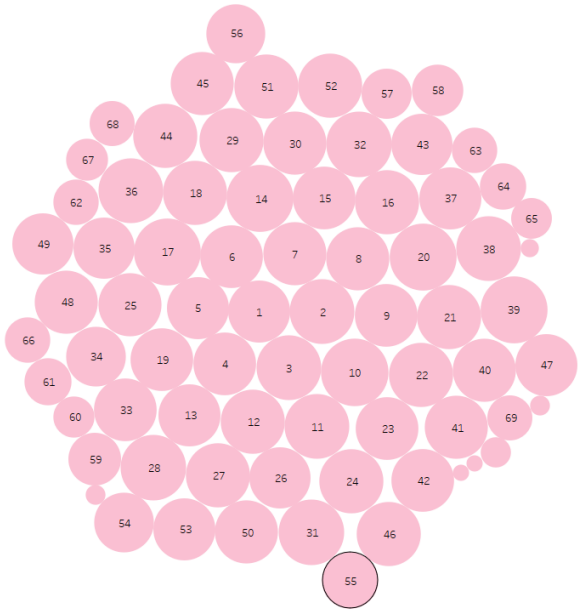
Design
I tried several ways to emphasize that the number 55 has been underutilized in the lottery drawings. I used a bubble plot because I thought it looked kind of lotto balls, and I circled the one I wanted to focus on because I thought it was important to draw attention there. In both the histogram and the horizontal bar chart, I used a contrasting color to highlight the specific bar that I wanted to focus. I really liked the horizontal bar chart with its pleasing sorted bars and just enough room to make out what each one was. I used the small multiples chart to get a comparison of various formats, and I was impressed at how effectively it draws attention to the point and compares the different formats. I'd never used a small multiples chart before, and I can see the utility. The highlighted scatter plot was a lot of work to learn and implement, but I also like the attention it draws to exactly how under-drawn the number 55 has been across games and formats. The text chart was perfect for the small amount of information I had because it's not too busy, and it clearly presents how small the change in the odds would be.

Ethical Considerations
I don't see any ethical issues with using this data. It's made available expressly for public information and use, provided by government entities without limitations.

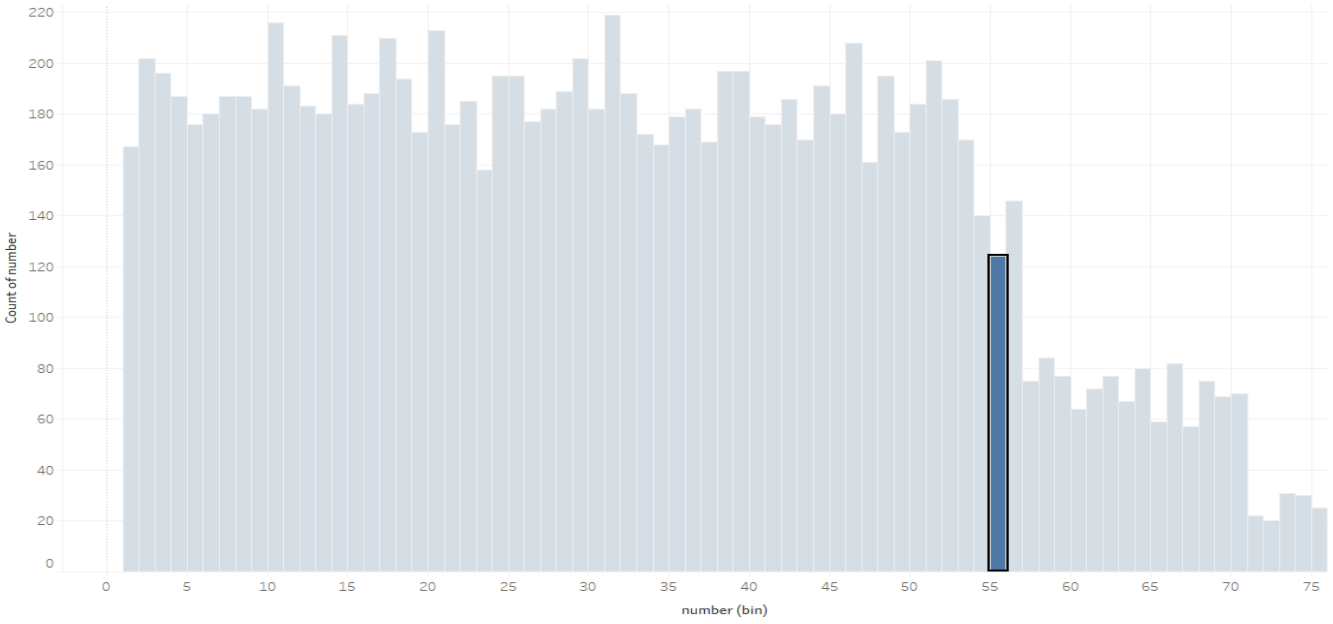
Dear Sirs and Madams of the MUSL and NYGC,

We, the undersigned, would like to express a desire for the number 55 to return to prominence in the hearts and minds of Americans by reformatting your Mega Millions, Pick 10, and Powerball lottery games to choose the winning balls from a pool of 55 balls.

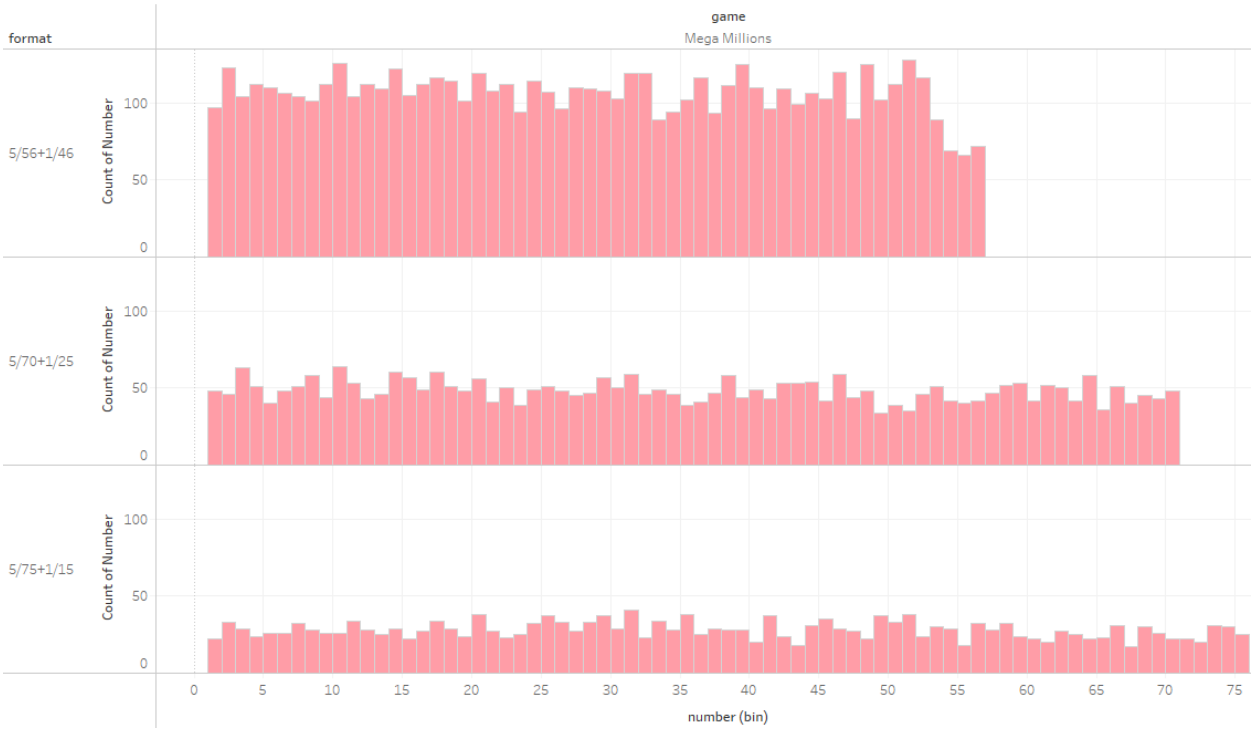
Whereas, the number 55 has not been adequately featured in the MUSL games over the years. In fact, among all the numbers drawn, the only numbers that show up less often than 55 are the numbers that have been added in reformats over the years; 55 is smaller than any other bubble in the original 56.



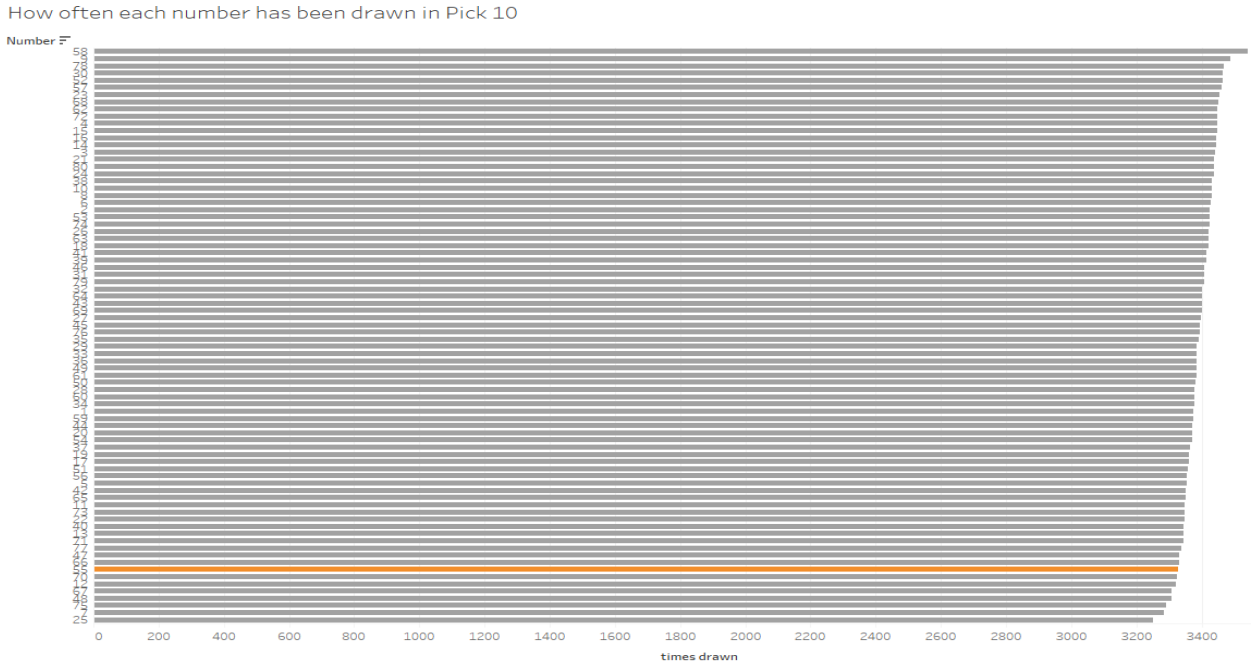
Whereas, the Mega Millions winning numbers over the last eleven years, show the disappointingly low occurrence of the number 55, again woefully behind all of the other original numbers.



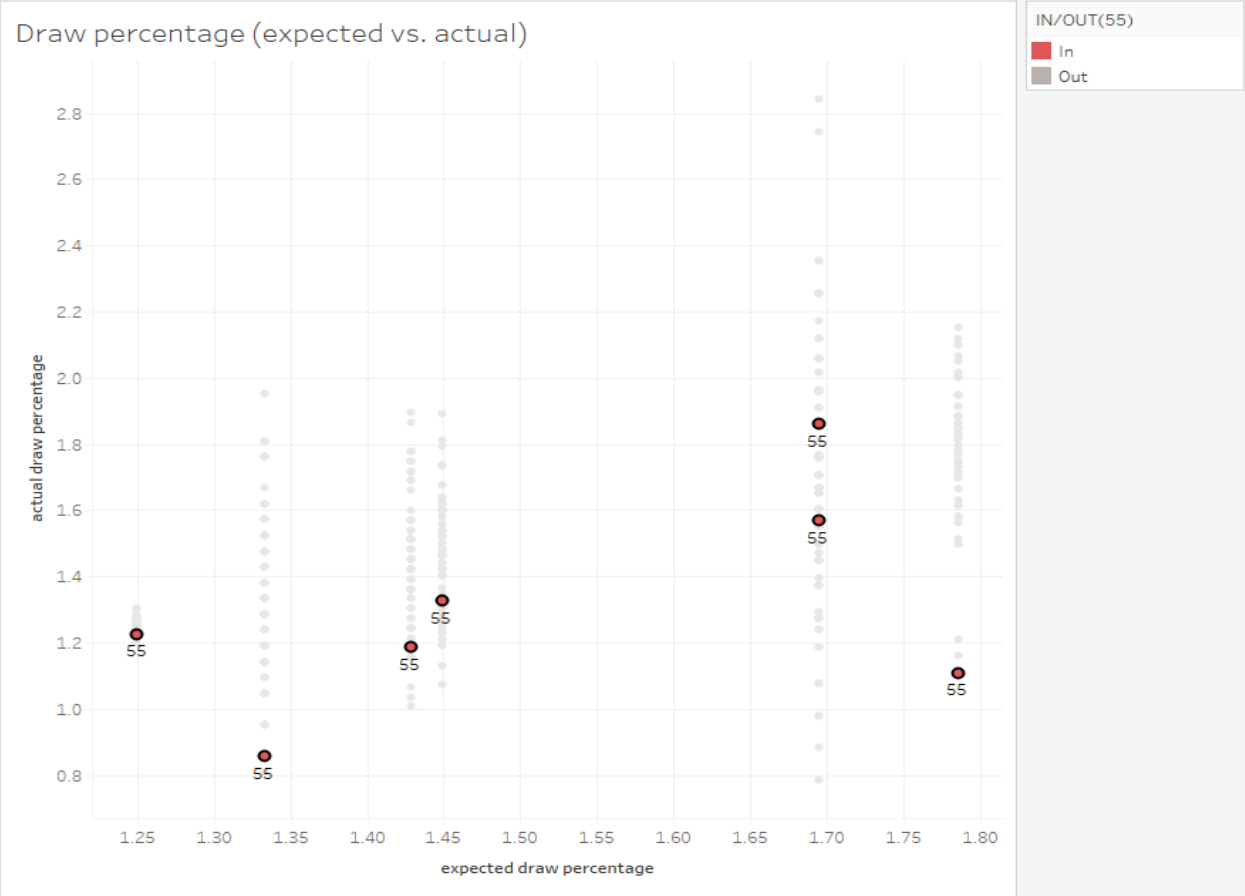
Whereas, the Mega Millions winning numbers through the last three formats, broken out by format, show how rarely 55 has been picked. It's suspiciously low in 5/56+1/46 format (along with 54 and 56), and suspiciously low again in the 5/75+1/15 format. I know that the sample sizes aren't huge, but it's really hard to believe that poor 55 is just that unlucky!



Whereas, even in Pick 10 game, which has had the most drawings and therefore has the largest sample size and the most likelihood of an even distribution of drawn numbers, the number 55 is in the bottom ten, well below the expected number of draws.



Whereas, plotting the expected draw percentage of each number versus the actual draw percentage represented by all the winning numbers to date in all formats of the Mega Millions, Pick 10, and Powerball games, the draw rate of the number 55 is well below its expected value. I've highlighted the number 55 in each of the formats of each game in order to impress upon you how underrepresented 55 has been, ranking in the bottom half of all the formats, near the bottom in most of them, and actually at the bottom for more than one! It's heartbreaking to see the number 55 holding such a disappointingly low draw rate across the formats.



Whereas, the lottery games should restor the prominence and dignity of the number 55, to establish it in relevance for the US people. If it's going to be so woefully under-drawn as part of winning numbers, it should be featured directly in the format of the games.

WE PROPOSE that the Mega Millions, Pick 10, and Powerball games all change formats in order to highlight the number 55. Such a reformatting could be done without a significant change in the odds of winning each game, as shown here:

Proposed Format Change and Effect on Odds

Game	Current format	Current odds	Proposed format	Proposed odds
Mega Millions	5/70+1/25	1 in 302,575,350	5/55+1/87	1 in 302,652,207
Pick 10	10/20/80	1 in 8,911,711	10/15/55	1 in 9,739,943
Powerball	5/69+1/26	1 in 292,201,338	5/55+1/84	1 in 292,215,924

Emphasizing the value and relevance of the number 55 is the right thing to do, in recognition of its extensive service as a milestone age and a ubiquitous speed limit. The following citizens, each living in a jurisdiction that sells tickets for at least one of the games in question (Mega Millions, Pick 10, and Powerball), do hereby petition the MUSL and the NYGC to modify that format of those games to incorporate the iconic 55 as the standard number of white balls in all of those games.

Respectfully submitted and first signed,

Christopher J. Kellogg

Kellogg640Week7and8

October 20, 2024

[1]:
```
## Chris Kellogg
## DSC640-T301
## Weeks 7 & 8

## ############################################################################
## Weeks 7 & 8 Exercises
## ############################################################################

##
## load necessary packages
##

import pandas as pd
```

[2]:
```
##
## read the given files
##

# read the data
csv_directory = 'LotteryNumbers/'
df_mm = pd.read_csv(
    csv_directory +
    'Lottery_Mega_Millions_Winning_Numbers__Beginning_2002_20240421.csv'
)
df_p10 = pd.read_csv(
    csv_directory +
    'Lottery_Pick_10_Winning_Numbers__Beginning_1987_20240421.csv'
)
df_pb = pd.read_csv(
    csv_directory +
    'Lottery_Powerball_Winning_Numbers__Beginning_2010_20240421.csv'
)
```

[3]:
```
##
## push all the numbers from a row into the list
##
def push_all(row):
```

```
    # for each number in the row
    for number in row['Winning Numbers'].split():

        # build a date key
        date_part = row['Draw Date'].split('/')
        date = ''.join((date_part[2], date_part[0], date_part[1]))

        # push the date key and number into the list
        data.append([int(date), number])
```

[4]:
```
##
## push all but one of the numbers from a row into the list
##
def push_all_but_one(row):

    number_list = row['Winning Numbers'].split()
    number_list.pop()

    # for each number in the row
    for number in number_list:

        # build a date key
        date_part = row['Draw Date'].split('/')
        date = ''.join((date_part[2], date_part[0], date_part[1]))

        # push the date key and number into the list
        data.append([int(date), number])
```

[5]:
```
##
## create a new dataframe with all numbers for each game
##

# Mega Millions
data = []
df_mm.apply(push_all, axis=1)
df_mm_n = pd.DataFrame(data, columns=['date', 'number'])

# Pick 10
data = []
df_p10.apply(push_all, axis=1)
df_p10_n = pd.DataFrame(data, columns=['date', 'number'])

# Powerball
data = []
df_pb.apply(push_all_but_one, axis=1)
df_pb_n = pd.DataFrame(data, columns=['date', 'number'])
```

[12]:
```
##
## create function for calculating expected number of draws
##
def calculate_expected_draws(df, white_balls):

    new_df = df \
        .groupby(['game','format','number']) \
        .agg(draws = ('date', 'count')) \
        .reset_index()
    total_draws = sum(new_df.draws)
    new_df['percentage'] = \
        new_df.draws / \
        total_draws \
        * 100.0

    new_df['white_balls'] = white_balls

    new_df['expected'] = 1 / white_balls * 100.0

    return new_df
```

[13]:
```
##
## create separate dataframes for each format of Mega Millions
##

# the 5/56+1/46 format ran until 2013-10-22
df_mm_n__5_56__1_46 = pd.DataFrame(df_mm_n.query('date < 20131022'))
df_mm_n__5_56__1_46['game'] = 'Mega Millions'
df_mm_n__5_56__1_46['format'] = '5/56+1/46'
df_mm_p__5_56__1_46 = calculate_expected_draws(df_mm_n__5_56__1_46, 56)

# the 5/75+1/15 format ran from 2013-10-22 until 2017-10-31
df_mm_n__5_75__1_15 = pd.DataFrame(df_mm_n.query('date >= 20131022 and date <␣
 ↪20171031'))
df_mm_n__5_75__1_15['game'] = 'Mega Millions'
df_mm_n__5_75__1_15['format'] = '5/75+1/15'
df_mm_p__5_75__1_15 = calculate_expected_draws(df_mm_n__5_75__1_15, 75)

# the 5/70+1/25 format has been running since 2017-10-31
df_mm_n__5_70__1_25 = pd.DataFrame(df_mm_n.query('date >= 20171031'))
df_mm_n__5_70__1_25['game'] = 'Mega Millions'
df_mm_n__5_70__1_25['format'] = '5/70+1/25'
df_mm_p__5_70__1_25 = calculate_expected_draws(df_mm_n__5_70__1_25, 70)
```

[14]:
```
##
## create separate dataframes for each format of Pick 10
##
```

```
# only one format of Pick 10 has run since its inception
df_p10_n['game'] = 'Pick 10'
df_p10_n['format'] = '10/20/80'
df_p10_p = calculate_expected_draws(df_p10_n, 80)
```

[15]:
```
##
## create separate dataframes for each format of Powerball
##

# the 5/59+1/39 format ran until 2012-01-15
df_pb_n__5_59__1_39 = pd.DataFrame(df_pb_n.query('date < 20120115'))
df_pb_n__5_59__1_39['game'] = 'Powerball'
df_pb_n__5_59__1_39['format'] = '5/59+1/39'
df_pb_p__5_59__1_39 = calculate_expected_draws(df_pb_n__5_59__1_39, 59)

# the 5/59+1/35 format ran from 2012-01-15 until 2015-10-07
df_pb_n__5_59__1_35 = pd.DataFrame(df_pb_n.query('date >= 20120115 and date <␣
 ↪20151007'))
df_pb_n__5_59__1_35['game'] = 'Powerball'
df_pb_n__5_59__1_35['format'] = '5/59+1/35'
df_pb_p__5_59__1_35 = calculate_expected_draws(df_pb_n__5_59__1_35, 59)

# the 5/59+1/39 format has been running since 2015-10-07
df_pb_n__5_69__1_26 = pd.DataFrame(df_pb_n.query('date >= 20151007'))
df_pb_n__5_69__1_26['game'] = 'Powerball'
df_pb_n__5_69__1_26['format'] = '5/59+1/39'
df_pb_p__5_69__1_26 = calculate_expected_draws(df_pb_n__5_69__1_26, 69)
```

[16]:
```
##
## write all the dataframes
##

# Mega Millions counts
df_mm_n__5_56__1_46.to_csv('LotteryNumbers/df_mm_n__5_56__1_46.csv',␣
 ↪index=False)
df_mm_n__5_75__1_15.to_csv('LotteryNumbers/df_mm_n__5_75__1_15.csv',␣
 ↪index=False)
df_mm_n__5_70__1_25.to_csv('LotteryNumbers/df_mm_n__5_70__1_25.csv',␣
 ↪index=False)

# Mega Millions percentages
df_mm_p__5_56__1_46.to_csv('LotteryNumbers/df_mm_p__5_56__1_46.csv',␣
 ↪index=False)
df_mm_p__5_75__1_15.to_csv('LotteryNumbers/df_mm_p__5_75__1_15.csv',␣
 ↪index=False)
```

```
df_mm_p__5_70__1_25.to_csv('LotteryNumbers/df_mm_p__5_70__1_25.csv',␣
 ↪index=False)

# Pick 10 counts
df_p10_n.to_csv('LotteryNumbers/df_p10_n.csv', index=False)

# Pick 10 percentages
df_p10_p.to_csv('LotteryNumbers/df_p10_p.csv', index=False)

# Powerball counts
df_pb_n__5_59__1_39.to_csv('LotteryNumbers/df_pb_n__5_59__1_39.csv',␣
 ↪index=False)
df_pb_n__5_59__1_35.to_csv('LotteryNumbers/df_pb_n__5_59__1_35.csv',␣
 ↪index=False)
df_pb_n__5_69__1_26.to_csv('LotteryNumbers/df_pb_n__5_69__1_26.csv',␣
 ↪index=False)

# Powerball percentages
df_pb_p__5_59__1_39.to_csv('LotteryNumbers/df_pb_p__5_59__1_39.csv',␣
 ↪index=False)
df_pb_p__5_59__1_35.to_csv('LotteryNumbers/df_pb_p__5_59__1_35.csv',␣
 ↪index=False)
df_pb_p__5_69__1_26.to_csv('LotteryNumbers/df_pb_p__5_69__1_26.csv',␣
 ↪index=False)
```