

QMBU450
Advanced Data Analysis With Python

Course Project – Final Report

Selin Nur Öztürk, 60160

Introduction

If you ask me to describe Turkey's situation in one word, I'd say 'good'. If you ask me to describe Turkey's situation in two words, I'd say 'not good'.

Süleyman Demirel
9th President of Turkey

In the previous semester, I took a course on Intercultural Management on which we discussed the different mindsets adopted by different generations in the recent history. It took my attention when the instructor pointed out that as the new generations, having less conservative, more questioning and more liberal mindsets compared to older generations, become eligible to vote in time, there will be a new wave of governance trend on all countries including Turkey, and it will be very different from today. So, I thought it would worth examining if that really holds by using the available data at hand and the methods that I have learned in this course.

Hypothesis

“Considering the hardship they have been going through since the early years of their life, as Generation-Z increasingly participates in Turkish elections, the ruling party (Justice and Development Party - AKP) becomes more likely to lose support among voters in the upcoming years’ elections, first of which is planned to be held in 2023.”

Methods

1. Data Collection & Feature Engineering

a. Elections

Although election systems don’t change very frequently in Turkey (at least for the recent years), the political parties tend to change their position from one election to another. For example, on 2014’s Presidential Election, Republican People’s Party (CHP) and Nationalist Movement Party (MHP) had a joint candidate, Ekmeleddin İhsanoğlu, competing against AKP’s candidate, Recep Tayyip Erdoğan. Only 4 years later, Recep Tayyip Erdoğan participated in the 2018 Presidential Election as the joint candidate of AKP and MHP, who have formed the electoral alliance named “Republic Alliance”. Another example is the position of the leftist parties who are mainly supported by the Kurdish voters: There were times that they were on the same side with the ruling party (especially during the “Kurdish Initiative” era). Nowadays, they seem to be an opposition to the ruling party. Thus, I tried to narrow it down to the election results that AKP had received only, not their allies, as it would complicate the data collection process.

Another challenge was the lack of datapoints. I could only find 11 election results from the official website, and almost half of them didn't have Generation-Z voters. The reason why I included the local elections and referendum results to create more datapoints is that there are only a couple of general elections that had been held since 2009.

I have gathered information regarding the previous elections held in Turkey from Higher Council of Elections (YSK)'s [website](#) that allowed querying for elections from 2009 until recently. The data gathered from this website included useful information for my research, such as the date of the election, which parties had received how many votes and the percentage of voters who have participated in that election.

I thought that the date of the election (month) can be affective on the election results because of a very generalized, personal observation: Voters who don't vote in favor of AKP tend to have a more secular lifestyle, be more educated and receive more income. Thus, as they consist of people who can afford holidays or a summer house, they don't simply vote when elections take place in summer. This would result in an increase of votes in favor of AKP. So, I saved the date information of the elections as well to check if that really affects the results. Again, from a personal intuition, I believe that voters who don't vote for AKP develop a tendency to protest the elections from time to time. When they don't vote, AKP gains more percentage. So, I also saved the participation rates from each election.

Using number of votes each party has received, I could determine the percentage of supporters of the ruling party among all the voters who had participated in the elections. I gathered the following results and exported them as a table.

2009 Local Elections (March)

I only considered the elections of metropolitan municipalities, as majority of the voters reside in metropolitans. On this election that 83.17% of the voters have participated in, AKP had received 42.19% of the votes.

2010 Referendum (September)

I considered the "Yes" voters as AKP supporters for this referendum. Only 77.2% of the voters participated on this election and the new constitution had a 57.86% acceptance rate.

2011 General Elections (June)

On this election that 87.24% of the voters have participated in, AKP had received 49.88% of the votes.

2014 Local Elections (March)

I only considered the elections of metropolitan municipalities, as majority of the voters reside in metropolitans. On this election that 89.53% of the voters have participated in, AKP had received 45.54% of the votes.

2014 Presidential Elections (August)

On this election that 77.05% of the voters have participated in, AKP's candidate, Recep Tayyip Erdoğan, had received 49.88% of the votes.

2015 General Elections (June)

On this election that 86.43% of the voters have participated in, AKP had received 40.66% of the votes, the lowest amount of votes in their history. This election is the first election that Generation-Z had participated in, which is the particular reason why I wanted to investigate the effect of Generation-Z voters on Turkish elections.

2015 General Elections (November)

AKP failed to form a coalition government with the other parties, so the election was held again. On this election that 87.45% of the voters have participated in, AKP had received 49.32% of the votes.

2017 Referendum (April)

I considered the "Yes" voters as AKP supporters for this referendum. 87.45% of the voters participated on this election and the constitution change had a 51.18% acceptance rate.

2018 Presidential Elections (June)

On this election that 88.19% of the voters have participated in, AKP's and MHP's joint candidate, Recep Tayyip Erdoğan, had received 52.37% of the votes.

2018 General Elections (June)

On this election that 88.18% of the voters have participated in, AKP had received 41.56% of the votes.

2019 Local Elections (March)

I only considered the elections of metropolitan municipalities, as majority of the voters reside in metropolitans. On this election that 84.51% of the voters have participated in, AKP had received 45.14% of the votes.

b. Population

I gathered the population data from TUIK's [website](#). The dataset included the number of people from each age between the years 2007 and 2020. Predicting the age groups in 2021, 2022 and 2023 was a challenge, and I implemented a method to overcome this.

While examining the dataset, I realized that the number of people from each age group depended on the previous years' data. For example, the number of 10 year-olds in 2007 was dependent on 2006's 9 year-olds, and the latter year usually had more population

than the previous year among the youngsters whereas it is the opposite for the older ages. I suppose the increase is due to immigration, and the decrease is due to the probability of dying out of natural causes among the senior citizens.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 2020 | 1081967 | 1179294 | 1250730 | 1294378 | 1315338 | 1333047 | 1349141 | 1296952 | 1295131 | 1252322 | 1264849 | 1271730 | 1312669 | 1304256 | 1266433 |
| 2019 | 1152226 | 1249728 | 1294179 | 1315616 | 1333387 | 1349589 | 1297385 | 1295822 | 1252767 | 1265476 | 1272307 | 1313236 | 1304694 | 1266910 | 1249023 |
| 2018 | 1215156 | 1287346 | 1309908 | 1328339 | 1344237 | 1292329 | 1290659 | 1248011 | 1260499 | 1267422 | 1308106 | 1300266 | 1262374 | 1244550 | 1225127 |
| 2017 | 1241005 | 1299435 | 1320128 | 1336237 | 1285044 | 1283268 | 1240928 | 1253807 | 1260812 | 1301954 | 1294443 | 1256788 | 1239524 | 1220347 | 1199768 |
| 2016 | 1257621 | 1312846 | 1330836 | 1280811 | 1277181 | 1237092 | 1250566 | 1258067 | 1299409 | 1292310 | 1254968 | 1237872 | 1218745 | 1198765 | 1218693 |
| 2015 | 1275773 | 1324357 | 1276046 | 1272573 | 1232767 | 1246025 | 1254290 | 1295986 | 1289256 | 1252162 | 1235240 | 1216139 | 1196409 | 1216790 | 1302407 |
| 2014 | 1282180 | 1271036 | 1268519 | 1229654 | 1243144 | 1250908 | 1293723 | 1287242 | 1250343 | 1233412 | 1214644 | 1194979 | 1215827 | 1302032 | 1324787 |
| 2013 | 1229012 | 1262391 | 1226023 | 1240578 | 1248411 | 1290772 | 1285713 | 1249085 | 1232158 | 1213506 | 1194004 | 1214945 | 1301549 | 1324578 | 1337089 |
| 2012 | 1215424 | 1219741 | 1236414 | 1245342 | 1282036 | 1283007 | 1245676 | 1231625 | 1207467 | 1191189 | 1213133 | 1300281 | 1318487 | 1339816 | 1327541 |
| 2011 | 1183833 | 1229002 | 1242948 | 1265286 | 1278755 | 1244302 | 1229298 | 1205437 | 1192662 | 1212447 | 1300589 | 1317884 | 1340660 | 1327518 | 1315954 |
| 2010 | 1163049 | 1230628 | 1272474 | 1273837 | 1238735 | 1225563 | 1204309 | 1189783 | 1212322 | 1299141 | 1317536 | 1340396 | 1328150 | 1316021 | 1266638 |
| 2009 | 1200635 | 1246778 | 1259743 | 1230724 | 1217441 | 1194415 | 1187422 | 1209319 | 1296541 | 1313950 | 1337945 | 1324707 | 1314193 | 1267969 | 1257552 |
| 2008 | 1155266 | 1228386 | 1219479 | 1200634 | 1194493 | 1176727 | 1199742 | 1293017 | 1308304 | 1340342 | 1325765 | 1318839 | 1271444 | 1256784 | 1299365 |
| 2007 | 1090011 | 1186129 | 1177723 | 1177092 | 1162951 | 1182909 | 1285105 | 1305104 | 1339149 | 1324560 | 1318083 | 1270789 | 1256545 | 1299119 | 1267122 |

Figure: The color pattern indicating the dependency between ages and years

Thus, I thought that it would make sense to investigate the differences between the year Y's X year-olds and year (Y-1)'s X-1 year-olds. I created another DataFrame to check these differences, as can be seen below. Each cell from age X and year Y shows the difference between the amount of X year-olds on year Y and the amount of (X-1) year-olds on year (Y-1).

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|------|---|--------|-------|-------|-------|-------|-------|------|------|-------|-------|-------|-------|-------|-------|-------|
| 2023 | 0 | 7794 | -1502 | 276 | -53 | 1409 | 182 | 1049 | 1415 | 1624 | 1809 | 1745 | 1767 | 1808 | 1314 | 1082 |
| 2022 | 0 | 11818 | -140 | 826 | 646 | 1736 | 688 | 1280 | 1498 | 1671 | 1772 | 1683 | 1676 | 1712 | 1218 | 965 |
| 2021 | 0 | 15842 | 1220 | 1377 | 1346 | 2063 | 1194 | 1512 | 1581 | 1717 | 1734 | 1622 | 1584 | 1617 | 1121 | 849 |
| 2020 | 0 | 27068 | 1002 | 199 | -278 | -340 | -448 | -433 | -691 | -445 | -627 | -577 | -567 | -438 | -477 | -851 |
| 2019 | 0 | 34572 | 6833 | 5708 | 5048 | 5352 | 5056 | 5163 | 4756 | 4977 | 4885 | 5130 | 4428 | 4536 | 4473 | 3689 |
| 2018 | 0 | 46341 | 10473 | 8211 | 8000 | 7285 | 7391 | 7083 | 6692 | 6610 | 6152 | 5823 | 5586 | 5026 | 4780 | 4328 |
| 2017 | 0 | 41814 | 7282 | 5401 | 4233 | 6087 | 3836 | 3241 | 2745 | 2545 | 2133 | 1820 | 1652 | 1602 | 1003 | 612 |
| 2016 | 0 | 37073 | 6479 | 4765 | 4608 | 4325 | 4541 | 3777 | 3423 | 3054 | 2806 | 2632 | 2606 | 2356 | 1903 | 1260 |
| 2015 | 0 | 42177 | 5010 | 4054 | 3113 | 2881 | 3382 | 2263 | 2014 | 1819 | 1828 | 1495 | 1430 | 963 | 375 | -453 |
| 2014 | 0 | 42024 | 6128 | 3631 | 2566 | 2497 | 2951 | 1529 | 1258 | 1254 | 1138 | 975 | 882 | 483 | 209 | -1475 |
| 2013 | 0 | 46967 | 6282 | 4164 | 3069 | 8736 | 2706 | 3409 | 533 | 6039 | 2815 | 1812 | 1268 | 6091 | -2727 | -1550 |
| 2012 | 0 | 35908 | 7412 | 2394 | 16750 | 4252 | 1374 | 2327 | 2030 | -1473 | 686 | -308 | 603 | -844 | 23 | -1638 |
| 2011 | 0 | 65953 | 12320 | -7188 | 4918 | 5567 | 3735 | 1128 | 2879 | 125 | 1448 | 348 | 264 | -632 | -67 | 1832 |
| 2010 | 0 | 29993 | 25696 | 14094 | 8011 | 8122 | 9894 | 2361 | 3003 | 2600 | 3586 | 2451 | 3443 | 1828 | -1331 | -534 |
| 2009 | 0 | 91512 | 31357 | 11245 | 16807 | -78 | 10695 | 9577 | 3524 | 5646 | -2397 | -1058 | -4646 | -3475 | 768 | -1869 |
| 2008 | 0 | 138375 | 33350 | 22911 | 17401 | 13776 | 16833 | 7912 | 3200 | 1193 | 1205 | 756 | 655 | 239 | 246 | -781 |
| 2007 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Figure: The differences between years and age groups

I used a Linear Regression model to estimate the differences on years 2021, 2022 and 2023 using the previous years' data for each age group (column), and added them to the table above. I made use of these differences to estimate the population distribution, and received the following dataset.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-------|
| 2023 | 1219773 | 1225059 | 1225072 | 1097945 | 1181287 | 1254162 | 1297642 | 1319138 | 1336936 | 1353775 | 1302013 | 1300365 | 1257506 | 1269955 | 1276340 | 13161 |
| 2022 | 1217265 | 1226574 | 1097669 | 1181340 | 1252753 | 1297460 | 1318089 | 1335521 | 1352151 | 1300204 | 1298620 | 1255739 | 1268147 | 1275026 | 1315504 | 1306 |
| 2021 | 1214756 | 1097809 | 1180514 | 1252107 | 1295724 | 1317401 | 1334241 | 1350653 | 1298533 | 1296848 | 1254056 | 1266471 | 1273314 | 1314286 | 1306377 | 1267 |
| 2020 | 1081967 | 1179294 | 1250730 | 1294378 | 1315338 | 1333047 | 1349141 | 1296952 | 1295131 | 1252322 | 1264849 | 1271730 | 1312669 | 1304256 | 1266433 | 1248 |
| 2019 | 1152226 | 1249728 | 1294179 | 1315616 | 1333387 | 1349589 | 1297385 | 1295822 | 1252767 | 1265476 | 1272307 | 1313236 | 1304694 | 1266910 | 1249023 | 1228 |
| 2018 | 1215156 | 1287346 | 1309908 | 1328339 | 1344237 | 1292329 | 1290659 | 1248011 | 1260499 | 1267422 | 1308106 | 1300266 | 1262374 | 1244550 | 1225127 | 1204 |
| 2017 | 1241005 | 1299435 | 1320128 | 1336237 | 1285044 | 1283268 | 1240928 | 1253807 | 1260812 | 1301954 | 1294443 | 1256788 | 1239524 | 1220347 | 1199768 | 1219 |
| 2016 | 1257621 | 1312846 | 1330836 | 1280811 | 1277181 | 1237092 | 1250566 | 1258067 | 1299409 | 1292310 | 1254968 | 1237872 | 1218745 | 1198765 | 1218693 | 1303 |
| 2015 | 1275773 | 1324357 | 1276046 | 1272573 | 1232767 | 1246025 | 1254290 | 1295986 | 1289256 | 1252162 | 1235240 | 1216139 | 1196409 | 1216790 | 1302407 | 1324 |
| 2014 | 1282180 | 1271036 | 1268519 | 1229654 | 1243144 | 1250908 | 1293723 | 1287242 | 1250343 | 1233412 | 1214644 | 1194979 | 1215827 | 1302032 | 1324787 | 1335 |
| 2013 | 1229012 | 1262391 | 1226023 | 1240578 | 1248411 | 1290772 | 1285713 | 1249085 | 1232158 | 1213506 | 1194004 | 1214945 | 1301549 | 1324578 | 1337089 | 1325 |
| 2012 | 1215424 | 1219741 | 1236414 | 1245342 | 1282036 | 1283007 | 1245676 | 1231625 | 1207467 | 1191189 | 1213133 | 1300281 | 1318487 | 1339816 | 1327541 | 1314 |
| 2011 | 1183833 | 1229002 | 1242948 | 1265286 | 1278755 | 1244302 | 1229298 | 1205437 | 1192662 | 1212447 | 1300589 | 1317884 | 1340660 | 1327518 | 1315954 | 1268 |
| 2010 | 1163049 | 1230628 | 1272474 | 1273837 | 1238735 | 1225563 | 1204309 | 1189783 | 1212322 | 1299141 | 1317536 | 1340396 | 1328150 | 1316021 | 1266638 | 1257 |
| 2009 | 1200635 | 1246778 | 1259743 | 1230724 | 1217441 | 1194415 | 1187422 | 1209319 | 1296541 | 1313950 | 1337945 | 1324707 | 1314193 | 1267969 | 1257552 | 1297 |
| 2008 | 1155266 | 1228386 | 1219479 | 1200634 | 1194493 | 1176727 | 1199742 | 1293017 | 1308304 | 1340342 | 1325765 | 1318839 | 1271444 | 1256784 | 1299365 | 1266 |
| 2007 | 1090011 | 1186129 | 1177723 | 1177092 | 1162951 | 1182909 | 1285105 | 1305104 | 1339149 | 1324560 | 1318083 | 1270789 | 1256545 | 1299119 | 1267122 | 1218 |

Figure: Age distributions, real (2007-2020) and estimated (2021-2023)

Next, I determined the voter population by eliminating the ages below 18 from the above table to find the population percentages belonging to each generation using the range of years they were born in. After summing up the people belonging in the same generation and dividing them by the total voter population on that year, I received the following table.

| | silent_gen | boomers | gen_x | gen_y | gen_z | gen_alpha |
|------|----------------------|---------------------|---------------------|---------------------|----------------------|-----------|
| 2023 | 0.048913449319890764 | 0.17085916337156898 | 0.274636968624884 | 0.32239657649910725 | 0.18319384218454898 | 0.0 |
| 2022 | 0.048374880070565285 | 0.18067305616985654 | 0.27858098071183823 | 0.32675751479740667 | 0.16561356825033327 | 0.0 |
| 2021 | 0.04783215348767347 | 0.19025113447922493 | 0.2826850264605927 | 0.3313099722592297 | 0.14792171331327925 | 0.0 |
| 2020 | 0.047370349866147646 | 0.19942294015784942 | 0.28700503855294385 | 0.33607806162966913 | 0.13012360979339 | 0.0 |
| 2019 | 0.05260669782121394 | 0.20445959243075593 | 0.29084654636081614 | 0.34078435223321785 | 0.11130281115399615 | 0.0 |
| 2018 | 0.05765479882186995 | 0.2101382180393633 | 0.29592146431505534 | 0.3456208387254233 | 0.09066468009828808 | 0.0 |
| 2017 | 0.06282773335106592 | 0.21574975171006344 | 0.30133543224234915 | 0.351072760470174 | 0.0690143222634751 | 0.0 |
| 2016 | 0.06828357754694611 | 0.22134077964777116 | 0.3067146459531966 | 0.3569676239247213 | 0.04669337292736486 | 0.0 |
| 2015 | 0.07407958816095186 | 0.2270574188071423 | 0.3121355917277799 | 0.36307915984805544 | 0.023648241456070543 | 0.0 |
| 2014 | 0.07991314283780962 | 0.23287337490996204 | 0.3180370196762072 | 0.36917646257602116 | 0.0 | 0.0 |
| 2013 | 0.08572337240406765 | 0.2382876376915871 | 0.3239431328833984 | 0.3520458570209469 | 0.0 | 0.0 |
| 2012 | 0.09202712145658816 | 0.24425179384383286 | 0.3295365409903918 | 0.33418454370918715 | 0.0 | 0.0 |
| 2011 | 0.09826502692897005 | 0.2500484187383997 | 0.3359733481423988 | 0.3157132061902314 | 0.0 | 0.0 |
| 2010 | 0.10441732860858093 | 0.25588210997347594 | 0.34254141989713166 | 0.29715914152081147 | 0.0 | 0.0 |
| 2009 | 0.11051577311182297 | 0.26177611353154334 | 0.3491820517401423 | 0.2785260616164914 | 0.0 | 0.0 |
| 2008 | 0.1169489900015484 | 0.26929604183480005 | 0.356274708990901 | 0.2574802591741441 | 0.0 | 0.0 |
| 2007 | 0.1285970350862329 | 0.2744045764901946 | 0.36121332732720807 | 0.23578506109636446 | 0.0 | 0.0 |

Figure: Generations' percentages among voters

For instance, this table estimates that around 18% of the voters in 2023 will be Gen-Z.

c. Unemployment Rate & GDP Per Capita

To add other dimensions to my model that is relevant to the problems that Gen-Z has been dealing with, I used a third party API, "quandl", to gather World Bank data about Turkey's unemployment rates and GDP per capita per year. I implemented a Linear Regression model to estimate the unemployment rates and GDP per capita based on previous years' data. However, I made use of only the recent years to preserve the trends: For example, I checked the last 5 years' GDP's to estimate the future years' GDP's since it is not expected to have an economic boom in the following years without a dramatic change in the economic policies so it will probably continue to decrease. Also when I used all the datapoints to train the model, I received very absurd GDP estimations (such as 2 Billion \$ per capita in 2023) so I decided to consider only the recent years for making an estimation. The same strategy holds for the estimation of the unemployment rates as well.

2. Model

To start developing my model, I initially joined all the tables that I have created. Then, I received the following table as my training set.

| | attendance_rate | akp_favor | month | silent_gen | boomers | gen_x | gen_y | gen_z | gen_alpha | gdp_per_capita | unemployment_rate |
|------|-----------------|-----------|-------|---------------------|---------------------|---------------------|---------------------|--------------|-----------|----------------|-------------------|
| 2009 | 0.83170 | 0.42190 | 3 | 0.11051577311182... | 0.26177611353154334 | 0.3491820517401423 | 0.2785260616164914 | 0.0 | 0.0 | 9103.70988 | 12.55000 |
| 2010 | 0.77200 | 0.57960 | 9 | 0.104417328608... | 0.25588210997347594 | 0.34254141989713166 | 0.29715914152081147 | 0.0 | 0.0 | 10742.43007 | 10.66000 |
| 2011 | 0.57240 | 0.49800 | 6 | 0.098265026928... | 0.2500484187383997 | 0.3359733481423988 | 0.3157132061902314 | 0.0 | 0.0 | 11420.77326 | 8.80000 |
| 2014 | 0.89530 | 0.45540 | 3 | 0.0799131428378... | 0.23287337490996204 | 0.3180370196762072 | 0.36917646257602116 | 0.0 | 0.0 | 12157.33796 | 9.88000 |
| 2014 | 0.77050 | 0.51650 | 8 | 0.0799131428378... | 0.23287337490996204 | 0.3180370196762072 | 0.36917646257602116 | 0.0 | 0.0 | 12157.33796 | 9.88000 |
| 2015 | 0.86430 | 0.40660 | 6 | 0.074079588160... | 0.2270574188071423 | 0.3121355917277799 | 0.36307915984805544 | 0.023648... | 0.0 | 11006.24974 | 10.24000 |
| 2015 | 0.87400 | 0.49320 | 11 | 0.074079588160... | 0.2270574188071423 | 0.3121355917277799 | 0.36307915984805544 | 0.023648... | 0.0 | 11006.24974 | 10.24000 |
| 2017 | 0.87450 | 0.51180 | 4 | 0.062827733351... | 0.21574975171006344 | 0.30133543224234915 | 0.351072760470174 | 0.0690143... | 0.0 | 10591.47437 | 10.82000 |
| 2018 | 0.88190 | 0.52370 | 6 | 0.057654798821... | 0.2101382180393633 | 0.29592146431505534 | 0.3456208387254233 | 0.090664... | 0.0 | 9455.59365 | 10.89000 |
| 2018 | 0.88180 | 0.41560 | 6 | 0.057654798821... | 0.2101382180393633 | 0.29592146431505534 | 0.3456208387254233 | 0.090664... | 0.0 | 9455.59365 | 10.89000 |
| 2019 | 0.84510 | 0.45140 | 3 | 0.052606697821... | 0.20445959243075593 | 0.29084654636081614 | 0.34078435223321785 | 0.1113028... | 0.0 | 9126.56135 | 13.67000 |

I decided to eliminate some columns for convenience. For example, I didn't include the fractions of Generation-Alpha as it was always 0, and no members of Gen-Alpha would

be voting in 2023 so it would be pointless to train the model with that column. Moreover, I decided to disregard the Silent Generation column because they had a very low population percentage all the time. To obtain independence among the features, I also eliminated the Gen-X and Gen-Y columns. So my feature dataset is reduced to “attendance_rate”, “akp_favor”, “month”, “boomers”, “gen_z”, “gdp_per_capita” and “unemployment_rate” columns, and “akp_favor” as the estimation column.

Since I had a very limited number of datapoints, I decided to implement a very simple model (LinearRegression from sklearn) to estimate the percentage of AKP supporters on 2023 elections.

Coefficients are:

| attendance_rate | month | boomers | gen_z | gdp_per_capita | unemployment_rate |
|-----------------|---------|---------|--------|----------------|-------------------|
| -0.17175 | 0.01059 | 6.87 | 3.9917 | 0.00008 | 0.00772 |

To check the integrity of my model, I made use of the residual sum of squares method to estimate my error and I received 0.009, which is pretty good considering that we are working with percentages (0.009 is equivalent to 0.9%).

Findings

- On an election that will be held in **March 2023** with an attendance rate of **85%**, my model estimated the AKP supporters to reach **45.92%**.
- On an election that will be held in **March 2023** with an attendance rate of **75%**, my model estimated the AKP supporters to reach **47.64%**.
- On an election that will be held in **June 2023** with an attendance rate of **85%**, my model estimated the AKP supporters to reach **49.10%**.
- On an election that will be held in **June 2023** with an attendance rate of **75%**, my model estimated the AKP supporters to reach **50.81%**.
- On an early election that will be held in **March 2022** with an attendance rate of **85%**, my model estimated the AKP supporters to reach **46.67%**.
- On an early election that will be held in **March 2022** with an attendance rate of **75%**, my model estimated the AKP supporters to reach **48.39%**.
- On an early election that will be held in **June 2022** with an attendance rate of **85%**, my model estimated the AKP supporters to reach **49.85%**.
- On an early election that will be held in **June 2022** with an attendance rate of **75%**, my model estimated the AKP supporters to reach **51.57%**.
- On an early election that would have been held in **March 2021** with an attendance rate of **85%**, my model estimated the AKP supporters to reach **47.59%**.

- On an early election that would have been held in **March 2021** with an attendance rate of **75%**, my model estimated the AKP supporters to reach **49.31%**.
- On an early election that would have been held in **June 2022** with an attendance rate of **85%**, my model estimated the AKP supporters to reach **50.77%**.
- On an early election that would have been held in **June 2022** with an attendance rate of **75%**, my model estimated the AKP supporters to reach **52.49%**.

To visualize these findings a little bit more clearly, let me put them on a table.

| | 75% Attendance Rate | | 85% Attendance Rate | |
|-------------|---------------------|--------|---------------------|--------|
| | March | June | March | June |
| 2021 | 49.31% | 52.49% | 47.59% | 50.77% |
| 2022 | 48.39% | 51.57% | 46.67% | 49.85% |
| 2023 | 47.64% | 50.81% | 45.92% | 49.10% |

Here are the estimated percentages of voters belonging to each generation in 2021, 2022 and 2023:

| | silent_gen | boomers | gen_x | gen_y | gen_z | gen_alpha |
|------|----------------------|---------------------|---------------------|---------------------|---------------------|-----------|
| 2023 | 0.048913449319890764 | 0.17085916337156898 | 0.274636968624884 | 0.32239657649910725 | 0.18319384218454898 | 0.0 |
| 2022 | 0.048374880070565285 | 0.18067305616985654 | 0.27858098071183823 | 0.32675751479740667 | 0.16561356825033327 | 0.0 |
| 2021 | 0.04783215348767347 | 0.19025113447922493 | 0.2826850264605927 | 0.3313099722592297 | 0.14792171331327925 | 0.0 |

As we can see, naturally, the percentage of Baby Boomers generation is decreasing as the percentage of Gen-Z is increasing from 2021 to 2023. In the table that shows AKP's percentage votes' estimations, there is a decreasing trend on each scenario from 2021 to 2023. Thus, we can say that my hypothesis is true: "As Generation-Z increasingly participates in Turkish elections, the ruling party (Justice and Development Party - AKP) becomes more likely to lose support among voters in the upcoming years' elections, first of which is planned to be held in 2023."

Another factor that affects AKP's votes are the month that the election takes place: When the election is held during summer, we can say that they are likely to receive more votes.

Finally, the participation rate among voters also affects AKP's votes. When more people decide to vote, AKP seems to be more likely to lose majority.

Final Words

As an engineering student with no background on social sciences, I am fully aware that election predictions cannot be made purely on quantitative data. There are many other factors (social, economic, global events...) that directly affect the election results, and it requires a higher level of analysis that includes both qualitative and quantitative evaluation. This project was my attempt to make an educated guess about the future elections that will be held in Turkey, and also I see it as an opportunity to implement what I have learned in the QMBU450 course throughout the semester.