Hypothesis for each question:

1. Are arrests in Bronx and Brooklyn proportional to their populations?
    a. H_0: Arrests in the Bronx and Brooklyn are proportional to their populations
    b. H_1: Arrests in the Bronx and Brooklyn are not proportional to their populations
2. Are there more arrests on Federal holidays?
    a. H_0: There is no difference in the number of arrests on federal holidays and non-federal holidays
    b. H_1: There is a difference in the number of arrests on federal holidays and non-federal holidays
3. Is there an association between gender and race (Black vs. Asian) in the population of Queens?
    a. H_0: Sex and race (Black/Asian) are independent. The proportion of males and females is the same across the racial groups.
    b. H_1: Sex and race (Black/Asian) are dependent. The proportion of males and females differs across the racial groups.

1. Are arrests in Bronx and Brooklyn proportional to their populations?
    a. H_0: Arrests in the Bronx and Brooklyn are proportional to their populations
    b. H_1: Arrests in the Bronx and Brooklyn are not proportional to their populations

Test run:

- Chi-squared test Goodness of fit
    - Tests on whether the observed distribution a single categorical variable (Number of arrests) matches an expected distribution

Assumptions:

1. Data is categorical:
    a. Chi-squared is for categorical data
    b. The Boroughs are categories, and the number of arrests and population are counts
2. Observations must be independent
    a. Each arrest should be independent of others. An arrest in Brooklyn should not influence an arrest in the Bronx
    b. Each arrest is recorded as a separate event, and an individual can only be arrested once per crime at a single location. However, external factors such as policing strategies, and related case crime rates may introduce some level of dependence between boroughs.
        i. Because only arrest count information is available the results should be interpreted cautiously, i.e. Not making policy decisions based on results
3. Counts should be sufficiently large
    a. The expected frequency for each category should be at least 5 to ensure that the Chi-squared approximation is valid.
    b. Expected arrests were much larger than 5 (63997 for Brooklyn and 33895 for Bronx)
4. The data is a Random Sample
    a. The arrest data covered the entirety of both boroughs so as not to be biased in collection.
5. No More Than 20% of Expected Frequencies Should Be Less Than 5
    a. There are only two categories (Brooklyn and Bronx) both are above 5

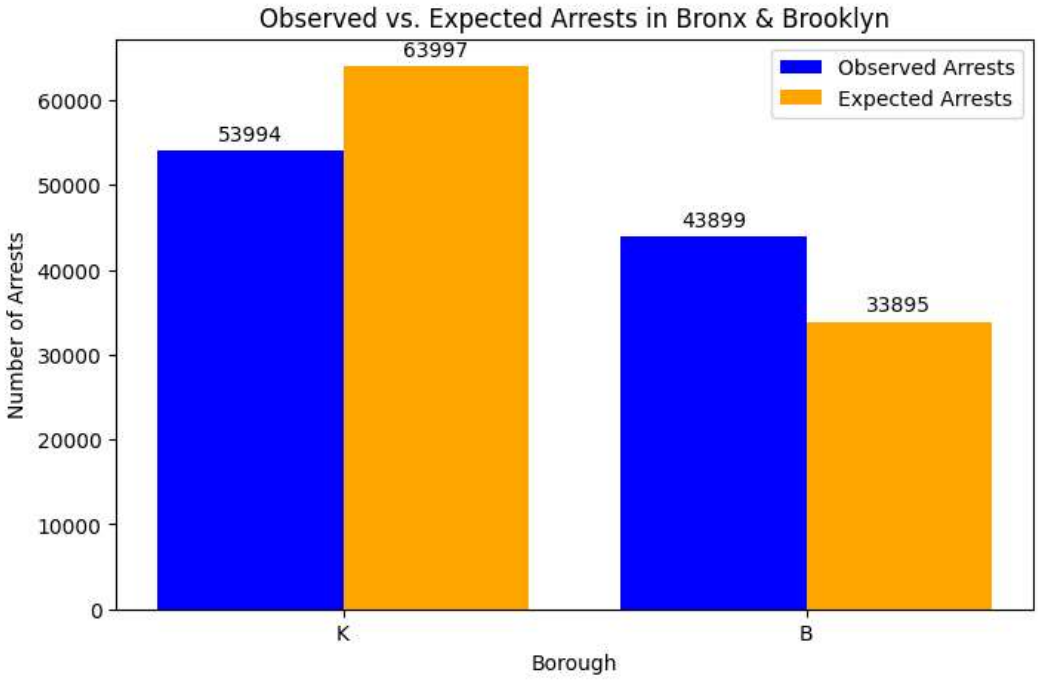
Alpha = .05

Chi-Square Statistic: 4515.636750396534

P-value: 0.0

P < alpha so **reject the null hypothesis** (H_1): Arrests are NOT proportional to population.

Visualization:



2. Are there more arrests on Federal holidays?
    a. H_0: There is no difference in the number of arrests on federal holidays and non-federal holidays
    b. H_1: There is a difference in the number of arrests on federal holidays and non-federal holidays

Test run:

- Two sample t-test
    - The two-sample t-test allows you to assess if there is a statistically significant difference between groups.

Results:

T-Statistic: -1.5537383022050255

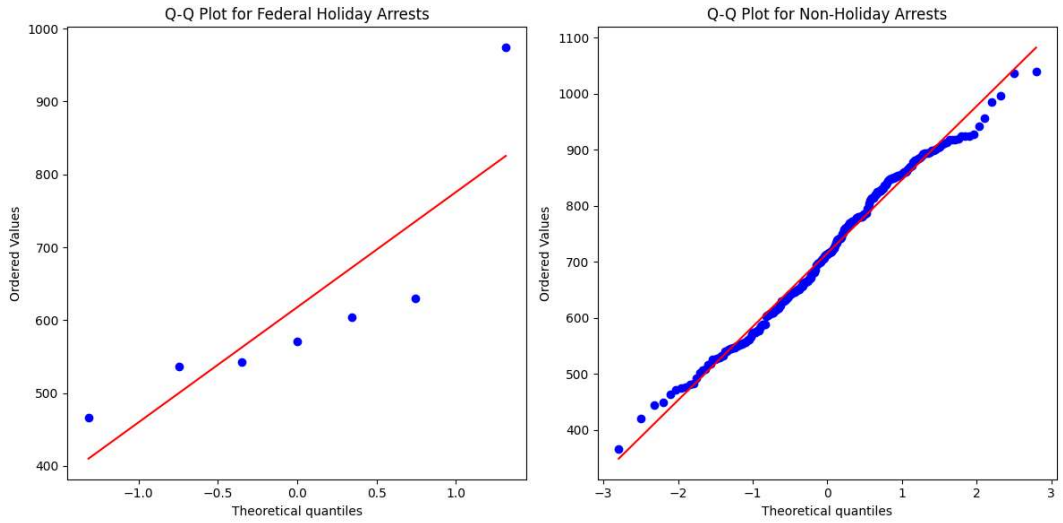
P-Value: 0.16966995710225058

Alpha: .05

P > alpha so we **fail to reject the null hypothesis** (H_0): No significant difference in the number of arrests on holidays vs. non-holidays.

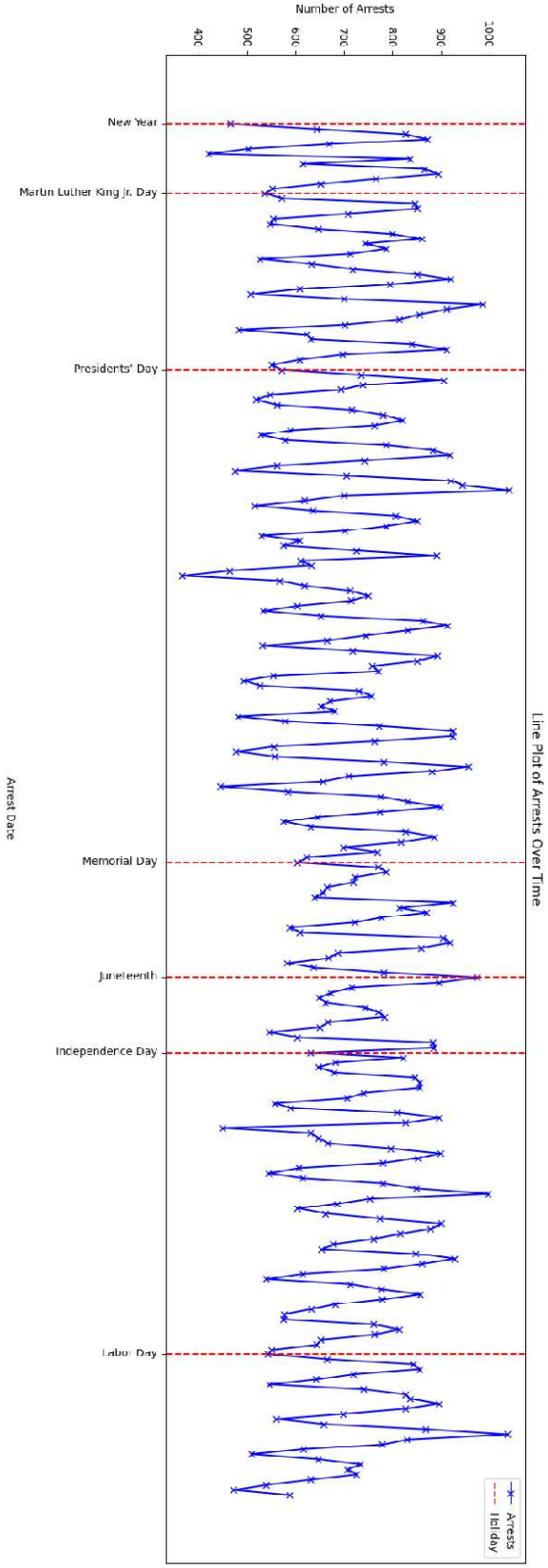
Assumptions:

1. The samples from Federal holidays and non-holidays are independent of each other. Arrests on consecutive days did not influence each other. The holidays are completely different days, and there were no repeated days
2. Both groups (holiday and non-holiday arrest counts) are approximately normally distributed:



Note: The QQ plot for federal holidays shows that the data does not really follow a normal distribution. This means that the results of this test are probably not very accurate – policy decisions shouldn't be made with the current test results.

Visualization:



3. Is there an association between gender and race (Black vs. Asian) in the population of Queens?
    a. H_0: Sex and race (Black/Asian) are independent. The proportion of males and females is the same across the racial groups.
    b. H_1: Sex and race (Black/Asian) are dependent. The proportion of males and females differs across the racial groups.

Test run:

- Chi-squared test of independence
    - Tests whether two categorical variables are independent of each other

Assumptions:

1. Independence of Observations
    a. Each arrest falls into one category; a person is either male or female, Black or Asian
    b. The information of one arrest does not influence another
        i. Though arrests of multiple perpetrators in a singular case might be present, this information is not present in the data. Therefore, independence **may be** violated. However, since only arrest count information is available the results of the test should be interpreted cautiously, i.e. not make any policy decisions based on the data.
2. Counts should be sufficiently large
    a. The expected frequency for each category should be at least 5 to ensure that the Chi-squared approximation is valid.
    b. Expected arrest were much larger than 5:

	Black	Asian/Pacific Islander
Male	12418.36	4595.64
Female	2762.64	1022.36

3. Random Sampling
    a. The data is taken from all over Queens borough from the 1st of January to the 30th of September 2024. Not one time, or precinct meaning the sample is random for all of Queens
4. Large sample size:
    a. The samples are all larger than 5, meaning they are sufficiently large.

Results:

Chi-Square Statistic: 5.2970480959256605

P-value: 0.02136159656502185

Degrees of Freedom: 1

Alpha: .05

P is less than alpha so:

Reject the null hypothesis! (H_1): Sex and race (Black/Asian) are dependent. The proportion of males and females differs across the racial groups.

Visualization: