Bay Area Bike Rental Operation Research

Data Analysis

# Introduction

This data analysis serves as the first step in developing a predictive model that will predict the number of bikes leaving each station and the number of bikes being returned to each station in the next three days. It is split up into two parts, one which focusses on 97.5% of the data, excluding the very longest trips such that the maximum duration is 85 minutes, while the other includes these extreme values to reflect the fact that individuals can subscribe to possess a single bike for months at a time.

# Data Cleaning & Remove Extreme Values

The data was examined. There was no missingness in the categorical variables with one notable exception: weather events. 1473 out of the 1825 events were not recorded, and it was unclear whether the weather was merely sunny, or events such as rain or fog were missing. To account for this, weather events were imputed using the mean visibility to determine fog, haze, and mist, and the precipitation to impute rainfall.

Fog: visibility less than 1 km (0.62 miles)

Mist: visibility between 1 km (0.62 miles) and 2 km (1.2 mi)

Haze from 2 km (1.2 miles) to 5 km (3.1 miles)

Temperatures at 14 F (-10°C) is too cold for the air to contain super-cooled moisture, and no fog, haze, or mist is possible. No temperatures recorded reached this level.

Rain: any amount of precipitation, even trace

Thunderstorm: hot temperatures, cloud cover, wind-speed, but highly variable, and thus could not be imputed.

50 events were able to be imputed, though this still left 1423 events out of 1825 as “not recorded”.

|  |  |
| --- | --- |
| A white text box with black text  Description automatically generated | A graph with a line and dots  Description automatically generated with medium confidence |
| A diagram of a bike trip duration  Description automatically generated | A diagram of a bike trip duration  Description automatically generated |

Figure 1. Sequential Cleaning of Data

Outliers and extreme values were removed from the dataset. Outliers were considered bikes that have trip durations longer than a month but were not taken out by subscribers. These may be bikes that were forgotten to be returned. Only one bike was an outlier under this metric and was taken out by a regular customer for 287839 minutes (~200 days long), while the next longest duration was only 12007 minutes (~8 days).

From there, extreme values were removed. These are values that while technically feasible, were not considered for this analysis. The upper 2.5% of trips were removed, leaving the longest trip to now be 85 minutes long. Finally, cancelled trips were considered to be any trips that lasted less than 3 minutes, and so were excluded leaving the cleaned dataset containing 312,062 trips.

Notably, the Duration variable was considered unreliable, as it recorded trips lasting longer than the start and end times would suggest. Therefore a new variable was constructed using the difference in time between the start and end time-dates.

# Bike & Station Maintenance

To determine which bikes and stations should be the focus of maintenance, it is necessary to understand which are used the most. One key component is determining “rush” hours.

A graph of a bike usage

Description automatically generated

Time of Day

Figure 2. Bar plot of the number of bike trips per the same 15-minute interval over 1 year

Rush hours are defined in this model as the busiest 25% of the day, the quarter of time when the most bikes are in concurrent use. The cutoff, the 75th percentile of trips per 15-minute interval, is 7968 trips over a single year.

In the Bay Area, there are two Rush Hour periods, one in the morning from 7:30-10:00 EST, and one in the afternoon from 15:45-19:15 EST. This is the time when the most bikes are being used, as well as which stations are being used during peak times. Rush-hour stations should be a priority for maintenance due to their importance, but likewise, should be repaired/maintained during off-hours. The top 10 starting stations and ending stations are listed in Table 1 and Table 2.

Table 1: Top 10 starting stations during rush hour

|  |  |
| --- | --- |
| **Starting Station** | **Number of Trips Per Year** |
| San Francisco Caltrain (Townsend at 4th) | 2649 |
| San Francisco Caltrain 2 (330 Townsend) | 838 |
| Temporary Transbay Terminal (Howard at Beale) | 798 |
| Grant Avenue at Columbus Avenue | 736 |
| Harry Bridges Plaza (Ferry Building) | 729 |
| San Jose Diridon Caltrain Station | 580 |
| Steuart at Market | 529 |
| Market at 10th | 453 |
| South Van Ness at Market | 371 |
| 2nd at Townsend | 310 |

Table 2: Top 10 ending stations during rush hour

|  |  |
| --- | --- |
| **Ending Station** | **Number of Trips Per Year** |
| San Francisco Caltrain (Townsend at 4th) | 1343 |
| Market at Sansome | 711 |
| 2nd at Townsend | 699 |
| Townsend at 7th | 645 |
| Embarcadero at Folsom | 525 |
| Temporary Transbay Terminal (Howard at Beale) | 514 |
| Steuart at Market | 491 |
| San Francisco Caltrain 2 (330 Townsend) | 480 |
| Harry Bridges Plaza (Ferry Building) | 427 |
| Market at 4th | 421 |

Likewise, weekend-station usage is important to note. The top 10 starting stations and ending stations are listed in Table 3 and Table 4.

Table 3: Top 10 starting stations during the weekends

|  |  |
| --- | --- |
| **Starting Station** | **Number of Trips Per Year** |
| Embarcadero at Sansome | 2868 |
| Harry Bridges Plaza (Ferry Building) | 2741 |
| Embarcadero at Bryant | 1488 |
| Market at 4th | 1477 |
| 2nd at Townsend | 1459 |
| Powell Street BART | 1341 |
| San Francisco Caltrain (Townsend at 4th) | 1296 |
| Grant Avenue at Columbus Avenue | 1191 |
| Market at 10th | 994 |
| Market at Sansome | 988 |

Table 4: Top 10 ending stations during the weekend

|  |  |
| --- | --- |
| **Ending Station** | **Number of Trips Per Year** |
| Embarcadero at Sansome | 2983 |
| Harry Bridges Plaza (Ferry Building) | 2852 |
| Market at 4th | 1690 |
| San Francisco Caltrain (Townsend at 4th) | 1581 |
| Powell Street BART | 1540 |
| 2nd at Townsend | 1520 |
| Embarcadero at Bryant | 1268 |
| Steuart at Market | 1116 |
| Townsend at 7th | 1026 |
| Civic Center BART (7th at Market) | 991 |

approximately 70% of the time there are some bikes in use. This is approximately constant across all 12 months (Figure 3.). Notably, this is a measure that includes at least a single bike on the road, though it

However, individual bikes are used, on average, approximately 1% of a month’s duration, or about 7.2 hours (Figure 4.). Bike usage is highest during the summer months.

A graph of a bike usage

Description automatically generated

Figure 3. Amount of time per month when at least a single bike is on the road

Bikes are used a majority of the time.

A graph of a bike usage

Description automatically generated

Figure 4. Average bike usage per month for individual bikes

Per month, the top 20 bikes used vary. Depending on the month that repairs are occurring, it may be reasonable to focus on the bikes that are going to be used in the upcoming months, that way they are in top shape for customers and subscribers. Likewise, knowing when each bike is in use can help for scheduling repairs so that preferential bikes are not unavailable. The top bikes are also likely to be in greater need for repair.

Table 5: Top 20 bikes (IDs listed) used per month

| **Rank** | **Jan.** | **Feb.** | **Mar.** | **Apr.** | **May** | **Jun.** | **Jul.** | **Aug.** | **Sept.** | **Oct.** | **Nov.** | **Dec.** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | 596 | 357 | 556 | 395 | 528 | 349 | 467 | 516 | 300 | 459 | 302 | 576 |
| **2** | 473 | 475 | 335 | 386 | 339 | 298 | 511 | 282 | 318 | 590 | 304 | 306 |
| **3** | 335 | 253 | 289 | 586 | 488 | 388 | 497 | 496 | 266 | 57 | 363 | 397 |
| **4** | 337 | 466 | 583 | 333 | 410 | 170 | 321 | 395 | 536 | 510 | 573 | 434 |
| **5** | 436 | 465 | 586 | 588 | 537 | 352 | 599 | 318 | 558 | 259 | 576 | 555 |
| **6** | 506 | 517 | 303 | 360 | 299 | 467 | 550 | 285 | 555 | 578 | 353 | 486 |
| **7** | 561 | 580 | 547 | 560 | 444 | 686 | 388 | 343 | 361 | 242 | 511 | 379 |
| **8** | 375 | 370 | 497 | 256 | 276 | 304 | 309 | 440 | 392 | 382 | 372 | 288 |
| **9** | 573 | 266 | 418 | 464 | 249 | 432 | 201 | 170 | 537 | 593 | 402 | 409 |
| **10** | 313 | 249 | 598 | 305 | 534 | 437 | 567 | 560 | 324 | 379 | 561 | 517 |
| **11** | 341 | 412 | 414 | 345 | 403 | 282 | 263 | 435 | 57 | 298 | 329 | 420 |
| **12** | 585 | 255 | 471 | 442 | 544 | 548 | 455 | 121 | 362 | 215 | 435 | 289 |
| **13** | 523 | 261 | 316 | 592 | 519 | 276 | 480 | 417 | 528 | 334 | 194 | 488 |
| **14** | 587 | 392 | 589 | 424 | 447 | 378 | 531 | 379 | 459 | 124 | 392 | 394 |
| **15** | 352 | 242 | 266 | 591 | 579 | 583 | 477 | 599 | 21 | 383 | 244 | 389 |
| **16** | 451 | 514 | 459 | 417 | 124 | 309 | 603 | 627 | 174 | 254 | 351 | 496 |
| **17** | 342 | 472 | 596 | 532 | 645 | 480 | 598 | 589 | 257 | 355 | 411 | 376 |
| **18** | 343 | 301 | 585 | 387 | 385 | 243 | 528 | 495 | 194 | 687 | 440 | 396 |
| **19** | 374 | 415 | 536 | 452 | 529 | 492 | 509 | 378 | 389 | 263 | 145 | 467 |
| **20** | 390 | 315 | 387 | 462 | 328 | 571 | 604 | 536 | 519 | 567 | 328 | 573 |

# Weather

When it came to examining the relationship between weather and trip information, focus was given to trip duration, city, and station.

Trip duration was found to be largely uncorrelated with the weather predictors.

A screenshot of a computer

Description automatically generated

Figure 5. Correlation coefficients for weather and trip variables

A graph with different colored squares

Description automatically generated

Figure 6. Correlation coefficients for weather and trip variables

When a linear regression model is constructed, several of the predictors are found to have significant beta-coefficients for modelling trip duration. Notably, even in a model with maximum weather predictors, it explains very little variation in the data. The model has a Multiple R-squared value of only 0.004418, and an Adjusted R-squared of only 0.004372.

Table 6: Linear Regression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t value | Significance |
| Intercept | 0. 060990 | 0.597314 | 0.102 | 0.919 |
| Max Temperature (F) | 0.138690 | 0.027996 | 4.954 | 7.28e-07 \*\*\* |
| Mean Temperature (F) | -0.321856 | 0.055776 | -5.770 | 7.91e-09 \*\*\* |
| Min Temperature (F) | 0.181230 | 0.028561 | 6.345 | 2.22e-10 \*\*\* |
| Max Visibility (miles) | 1.199726 | 0.057341 | 20.923 | < 2e-16 \*\*\* |
| Mean Visibility (miles) | -0.060178 | 0.032240 | -1.867 | 0.062 |
| Min Visibility (miles) | 0.099871 | 0.010893 | 9.168 | < 2e-16 \*\*\* |
| Max Wind Speed (mph) | -0.041831 | 0.005433 | -7.699 | 1.37e-14 \*\*\* |
| Mean Wind Speed (mph) | 0.107521 | 0.009261 | 11.610 | < 2e-16 \*\*\* |
| Max Gust Speed (mph) | -0.012178 | 0.002330 | -5.227 | 1.73e-07 \*\*\* |
| Precipitation (inches) | -0.106472 | 0.126715 | -0.840 | 0.401 |
| Cloud Cover | -0.125351 | 0.011834 | -10.592 | < 2e-16 \*\*\* |
| Events: Fog-Rain | 0.165805 | 0.151739 | 1.093 | 0.275 |
| Events: Not Recorded | -0.645434 | 0.075599 | -8.538 | < 2e-16 \*\*\* |
| Events: Rain | -0.850895 | 0.086758 | -9.808 | < 2e-16 \*\*\* |

Trip Duration ~ Max Temperature (F) + Mean Temperature (F) + Min Temperature (F) + Max Visibility Miles + Mean Visibility Miles + Min Visibility (miles) + Max Wind Speed (mph) + Mean Wind Speed (mph) + Max Gust Speed (mph) + Precipitation (inches) + Cloud Cover + Events

# Data Cleaning & Focus on All Possible Data

|  |  |
| --- | --- |
| A white text box with black text  Description automatically generated |  |

Figure 7. Sequential data cleaning

The only difference between the two analyses was that all possible values were considered. It is conceivable that a customer could rent a bike for up to 8 days, and so the only outlier was with a trip duration longer than a month but having not been taken out by subscribers. Like the first analysis, cancelled trips were considered to be any trips that lasted less than 3 minutes, and so were excluded leaving the cleaned dataset containing 320,283 trips.

# Bike & Station Maintenance

To determine which bikes and stations should be the focus of maintenance, it is necessary to understand which are used the most. One key component is determining “rush” hours.

A graph of red and white bars

Description automatically generated

Time of Day

Figure 8. Bar plot of the number of bike trips per the same 15-minute interval over 1 year

Rush hours are defined in this model as the busiest 25% of the day, the quarter of time when the most bikes are in concurrent use. The cutoff, the 75th percentile of trips per 15-minute interval, is 10236 trips over a single year.

In the Bay Area, there are three Rush Hour periods, one in the morning from 7:45-10:00 EST, one in the middle of the day from 12:45-13:00 EST, and one in the afternoon from 15:45-19:15 EST. This is the time when the most bikes are being used, as well as which stations are being used during peak times. Rush-hour stations should be a priority for maintenance due to their importance, but likewise, should be repaired/maintained during off-hours. The top 10 starting stations and ending stations are listed in Table 7 and Table 8.

Table 7: Top 10 starting stations for rush hour

|  |  |
| --- | --- |
| **Starting Station** | **Number of Trips Per Year** |
| San Francisco Caltrain (Townsend at 4th) | 2659 |
| San Francisco Caltrain 2 (330 Townsend) | 838 |
| Temporary Transbay Terminal (Howard at Beale) | 803 |
| Grant Avenue at Columbus Avenue | 728 |
| Harry Bridges Plaza (Ferry Building) | 724 |
| San Jose Diridon Caltrain Station | 581 |
| Steuart at Market | 531 |
| Market at 10th | 452 |
| South Van Ness at Market | 360 |
| 2nd at Townsend | 305 |

Table 8: top 10 ending stations for rush hour

|  |  |
| --- | --- |
| **Ending Station** | **Number of Trips Per Year** |
| San Francisco Caltrain (Townsend at 4th) | 1329 |
| Market at Sansome | 719 |
| 2nd at Townsend | 699 |
| Townsend at 7th | 645 |
| Embarcadero at Folsom | 521 |
| Temporary Transbay Terminal (Howard at Beale) | 511 |
| Steuart at Market | 487 |
| San Francisco Caltrain 2 (330 Townsend) | 476 |
| Market at 4th | 419 |
| Harry Bridges Plaza (Ferry Building) | 393 |

The top 10 ending stations are the same for the expanded data set as the data that excludes trips greater than the 97.5th percentile.

Likewise, weekend-station usage is important to note. The top 10 starting stations and ending stations are listed in Table 9 and Table 10.

Table 9: top 10 starting stations for weekend

|  |  |
| --- | --- |
| **Starting Station** | **Number of Trips Per Year** |
| Harry Bridges Plaza (Ferry Building) | 3154 |
| Embarcadero at Sansome | 3118 |
| Market at 4th | 1656 |
| Embarcadero at Bryant | 1568 |
| 2nd at Townsend | 1537 |
| Powell Street BART | 1473 |
| San Francisco Caltrain (Townsend at 4th) | 1359 |
| Grant Avenue at Columbus Avenue | 1296 |
| Powell at Post (Union Square) | 1088 |
| Market at Sansome | 1075 |

Like the rush hour stations, including the longest 2.5% of trips does not drastically change the top 10 starting stations for the weekend. One notable change is that Powell at Post (Union Square) is now the 9th most common station instead of Market at 10th, and given the proximity in the 1st and 2nd stations, as well as the 3rd and 4th stations, the additional trips changed the order (Table 3.)

Table 10: Top 10 ending stations for weekend

|  |  |
| --- | --- |
| **Ending Station** | **Number of Trips Per Year** |
| Embarcadero at Sansome | 3373 |
| Harry Bridges Plaza (Ferry Building) | 3163 |
| Market at 4th | 1869 |
| Powell Street BART | 1672 |
| San Francisco Caltrain (Townsend at 4th) | 1661 |
| 2nd at Townsend | 1580 |
| Embarcadero at Bryant | 1335 |
| Steuart at Market | 1219 |
| Grant Avenue at Columbus Avenue | 1098 |
| Market at Sansome | 1078 |

By including the longest 2.5% of trips, the 9th and 10th most common ending stations are changed. Though the number of trips the top 10 stations accrue annually is approximately equivalent indicating that the longest trips are not distributed in a way that biases towards one station or another.

approximately 70% of the time there are some bikes in use. This is approximately constant across all 12 months (Figure 2.). Notably, this is a measure that can be biased by at least a single bike on the road. When examining average bike use, individual bikes are used for approximately 1% of a month’s duration, or about 7.2 hours (Figure 3.). Bike usage is highest during the summer months.

A graph of a bike usage

Description automatically generated

Figure 9. Time per month when at least a single bike is used

Bikes are used the majority of the time.

A graph of a bike usage

Description automatically generated

Figure 10. Average time for individual bike use

Under this expanded model, the top 20 bikes used per month are different.

Table 11: Top 20 bikes (IDs listed) used per month

| **Rank** | **Jan.** | **Feb.** | **Mar.** | **Apr.** | **May** | **Jun.** | **Jul.** | **Aug.** | **Sept.** | **Oct.** | **Nov.** | **Dec.** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 661 | 285 | 4 | 342 | 638 | 601 | 228 | 80 | 653 | 660 | 582 | 355 |
| 2 | 509 | 294 | 206 | 602 | 118 | 164 | 5 | 438 | 604 | 388 | 410 | 541 |
| 3 | 314 | 630 | 415 | 542 | 377 | 134 | 454 | 579 | 175 | 455 | 325 | 452 |
| 4 | 61 | 76 | 662 | 513 | 482 | 382 | 113 | 676 | 422 | 228 | 424 | 75 |
| 5 | 55 | 419 | 514 | 239 | 370 | 85 | 251 | 309 | 670 | 474 | 563 | 250 |
| 6 | 553 | 543 | 233 | 531 | 495 | 124 | 598 | 319 | 15 | 531 | 653 | 344 |
| 7 | 657 | 243 | 357 | 495 | 345 | 394 | 395 | 343 | 174 | 265 | 416 | 571 |
| 8 | 534 | 52 | 566 | 256 | 421 | 654 | 297 | 333 | 188 | 323 | 586 | 550 |
| 9 | 363 | 386 | 33 | 318 | 643 | 90 | 524 | 243 | 21 | 644 | 488 | 379 |
| 10 | 561 | 226 | 442 | 208 | 351 | 33 | 572 | 172 | 393 | 316 | 539 | 414 |
| 11 | 517 | 350 | 219 | 450 | 398 | 289 | 346 | 598 | 304 | 372 | 487 | 570 |
| 12 | 449 | 452 | 332 | 357 | 286 | 348 | 548 | 538 | 489 | 436 | 451 | 467 |
| 13 | 345 | 584 | 450 | 582 | 30 | 352 | 467 | 308 | 76 | 67 | 343 | 476 |
| 14 | 596 | 591 | 354 | 126 | 598 | 531 | 398 | 8 | 518 | 72 | 420 | 390 |
| 15 | 246 | 552 | 552 | 337 | 593 | 281 | 252 | 586 | 368 | 521 | 246 | 330 |
| 16 | 533 | 307 | 600 | 500 | 335 | 431 | 416 | 354 | 370 | 347 | 379 | 448 |
| 17 | 343 | 418 | 525 | 474 | 575 | 373 | 600 | 327 | 459 | 600 | 523 | 435 |
| 18 | 291 | 425 | 476 | 293 | 521 | 380 | 497 | 582 | 583 | 627 | 99 | 558 |
| 19 | 301 | 518 | 437 | 397 | 310 | 539 | 594 | 541 | 496 | 76 | 414 | 508 |
| 20 | 324 | 476 | 560 | 549 | 379 | 598 | 511 | 384 | 392 | 567 | 427 | 363 |

# Weather

When examining the effect that weather has on bike transit, once more it is found that there is little correlation between weather and trip duration.

A screenshot of a graph

Description automatically generated

Figure 11: correlation coefficients for trip and weather variables

One interesting correlation is with the city of Palo Alto and maximum visibility. Trips occurring in Palo Alto strongly correlate with better visibility. However, given the lack of any other correlations of this magnitude with any other city, this is likely an artifact of the extended dataset, rather than indicating any significance in the behaviour of the people of Palo Alto.

A diagram with many different colored squares

Description automatically generated

Figure 12: Correlation coefficients for weather and trip variables

Formula: Trip Duration ~ Max Temperature (F) + Mean Temperature (F) + Min Temperature (F) + Max Visibility Miles + Mean Visibility Miles + Min Visibility (miles) + Max Wind Speed (mph) + Mean Wind Speed (mph) + Max Gust Speed (mph) + Precipitation (inches) + Cloud Cover + Events

Table 12: Linear regression results for a large number of weather predictors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t value | Significance |
| Intercept | -42.2799 | 34.3969 | -1.229 | 0.2190 |
| Max Temperature (F) | -0.7291 | 1.7252 | -0.423 | 0.6726 |
| Mean Temperature (F) | 0.5231 | 3.4364 | 0.152 | 0.8790 |
| Min Temperature (F) | 0.3395 | 1.7598 | 0.193 | 0.8470 |
| Max Visibility (miles) | 0.5538 | 3.2891 | 0.168 | 0.8663 |
| Mean Visibility (miles) | 8.7204 | 1.9949 | 4.371 | 1.24e-05 \*\*\* |
| Min Visibility (miles) | -0.3140 | 0.6744 | -0.466 | 0.6415 |
| Max Wind Speed (mph) | -0.3333 | 0.3353 | -0.994 | 0.3202 |
| Mean Wind Speed (mph) | -0.1458 | 0.5719 | -0.255 | 0.7987 |
| Max Gust Speed (mph) | -0.1493 | 0.1441 | -1.036 | 0.3002 |
| Precipitation (inches) | 19.3408 | 7.8567 | 2.462 | 0.0138 \* |
| Cloud Cover | -1.5120 | 0.7315 | -2.067 | 0.0387 \* |
| Events: Fog-Rain | 85.6192 | 9.4197 | 9.089 | < 2e-16 \*\*\* |
| Events: Not Recorded | -11.1001 | 4.6678 | -2.378 | 0.0174 \* |
| Events: Rain | -10.6972 | 5.3672 | -1.993 | 0.0463 \* |

When the model is refined to just include the statistically significant predictors, the R-squared values do not diminish substantially. The Multiple R-squared is now 0.0003843, compared to 0.0003964, and the Adjusted R-squared is 0.0003687 compared to 0.0003515. Overall, they remain quite poor.

Formula: Trip Duration ~ Mean Visibility Miles + Precipitation (inches) + Events

Table 14: Linear regression results for a reduced number of weather predictors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t value | Significance |
| Intercept | -48.319 | 10.728 | -4.504 | 6.67e-06 \*\*\* |
| Mean Visibility (miles) | 7.980 | 1.224 | 6.522 | 6.97e-11 \*\*\* |
| Precipitation (inches) | 15.892 | 7.456 | 2.131 | 0.03305 \* |
| Events: Fog-Rain | 82.720 | 9.084 | 9.106 | < 2e-16 \*\*\* |
| Events: Not Recorded | -11.887 | 4.415 | -2.692 | 0.00710 \*\* |
| Events: Rain | -12.860 | 4.918 | -2.615 | 0.00893 \*\* |

# Next Steps

A graph of a station coordinates

Description automatically generated

Figure 13. Station coordinates

One possible next step would be to aid in the imputation of weather results. Plotting the different latitudes and longitudes of each station show that several are clustered together. It is conceivable that the weather events for missing stations can be imputed by looking at adjacent stations.

# Appendix 1

## Correlation of Stations and Weather for 97.5% of the data

A close up of a text

Description automatically generated with medium confidence

A close up of words

Description automatically generated

A close up of a text

Description automatically generated

A close up of a text

Description automatically generated with medium confidence

A close-up of a map

Description automatically generated

A black and white text on a white background

Description automatically generated

A close up of words

Description automatically generated

A white background with black text

Description automatically generated

# Appendix 2

## Correlation of Stations and Weather for all conceivable data

A close up of a text

Description automatically generated

A close-up of a text

Description automatically generated

A close-up of a graph

Description automatically generated

A close up of a computer screen

Description automatically generated

A close up of words

Description automatically generated

A close-up of a flag

Description automatically generated

A graph of weather and trip variable correlation

Description automatically generated

A close-up of a graph

Description automatically generated