**APPENDIX D – EXPLORATORY DATA ANALYSIS**

**Module Group: Group 2**

**Project Group Number: Team 1**

**Team Name: WinxClub**

**Description of Business**

• Problem Statement / Hypothesis

“Dynamic ticket prices are mainly determined by date of flight”

The objective of this study is to investigate the extent to which dynamic ticket prices for American airline flights are determined by the date of the flight.

This will be done by analyzing historical ticket price data and considering various factors that influence pricing dynamics.

The findings will contribute to a better understanding of the factors influencing dynamic pricing for American airlines, and provide insights for American airlines and customers regarding pricing strategies, booking decisions, and market competitiveness.

**Details of Datasets**

• Scope of datasets

**Included**

* [Download page (bts.gov)](https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FHK&QO_fu146_anzr=b4vtv0%20n0q%20Qr56v0n6v10%20f748rB) - The dataset contains a wide range of data about different aspects of a flight. This comprehensive dataset allows for in-depth analysis and exploration of different factors affecting flights of different airlines.
* [Download page (bts.gov)](https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FHK&QO_fu146_anzr=b4vtv0%20n0q%20Qr56v0n6v10%20f748rB) - This dataset is a reference table for the previous dataset. It contains the full name of airlines, where the previous dataset only contains the airline code.
* [US Holiday Dates (2004 - 2021) | Kaggle](https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021) - This dataset contains dates of public holidays celebrated in the US

**Excluded**

* [Airline Safety | Kaggle](https://www.kaggle.com/datasets/mysarahmadbhat/airline-safety) - I excluded this dataset as the data is too old and may not be accurate. Although there is insightful data, airline safety measures have developed in the past few years. Hence, drawing conclusions from a 2014 dataset may not be the most context accurate.
* [Flight Bookingdataset | Kaggle](https://www.kaggle.com/datasets/akritiupadhyayks/flight-bookingdataset) - This dataset did not include column definitions. Hence some columns are undecipherable. For example, the ‘price’ column ranges from 1,100 to 123,000, which does not make sense for the average ticket price. There is also no clarification on whether the price includes taxes and other fees or if it is a two-way trip.

• Common or individual datasets

The dataset contains flight data from 2017-2019. It contains fields related to itinerary, market, airports, carriers, fares, distances, and passenger information. All flights listed are one way flights.

[Download page (bts.gov)](https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FHK&QO_fu146_anzr=b4vtv0%20n0q%20Qr56v0n6v10%20f748rB) (CSV format)

The dataset is shared with Parik, where he will be using the number of passengers column.

The columns i will be using are

* Year
* Quarter
* Airport Group
* Airline Name
* OriginAirportID
* OriginState
* DestAirportID
* DestState
* FarePerMile
* MktFare
* MktMilesFlown

The dataset contains US holiday data from 2004-2021. It contains fields related to date, day of the week and holiday name

[US Holiday Dates (2004 - 2021) | Kaggle](https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021) (CSV format). This dataset is only used by me.

The columns i will be using are

* Date
* Holiday
* Weekday
* Month
* Day
* Year

• Individual contribution *(Indicate Team Member name)*

o List datasets being used

*(as specified in the previous section under Common or individual datasets)*

* [Download page (bts.gov)](https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FHK&QO_fu146_anzr=b4vtv0%20n0q%20Qr56v0n6v10%20f748rB) Origin and Destination survey 2017 - 2019
* [Download page (bts.gov)](https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FHK&QO_fu146_anzr=b4vtv0%20n0q%20Qr56v0n6v10%20f748rB) Carriers reference table
* [US Holiday Dates (2004 - 2021) | Kaggle](https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021) US Holidays reference table

o Describe the data cleaning and transformation done

1. Filter dataset to only include columns relevant for price analysis
2. Create a new column ‘stops’ (count of stops in the flight) by counting the number of values in airport group column values.
3. Inner join the 2 tables to add the full airline name to the dataset
4. Create a new column for flight fare per mile by dividing the flight fare by the miles traveled
5. In US holidays dataset, group by year and quarter to get count of holidays for every quarter
6. Join the transformed US holidays dataset with the main dataset
7. Remove rows that have outliers in fare per mile e.g. fare per mile > 3 USD

o How is the dataset related to problem / hypothesis?

Detailed information on flight fares over time can be used to develop more accurate pricing models and inform users about best times to book tickets.

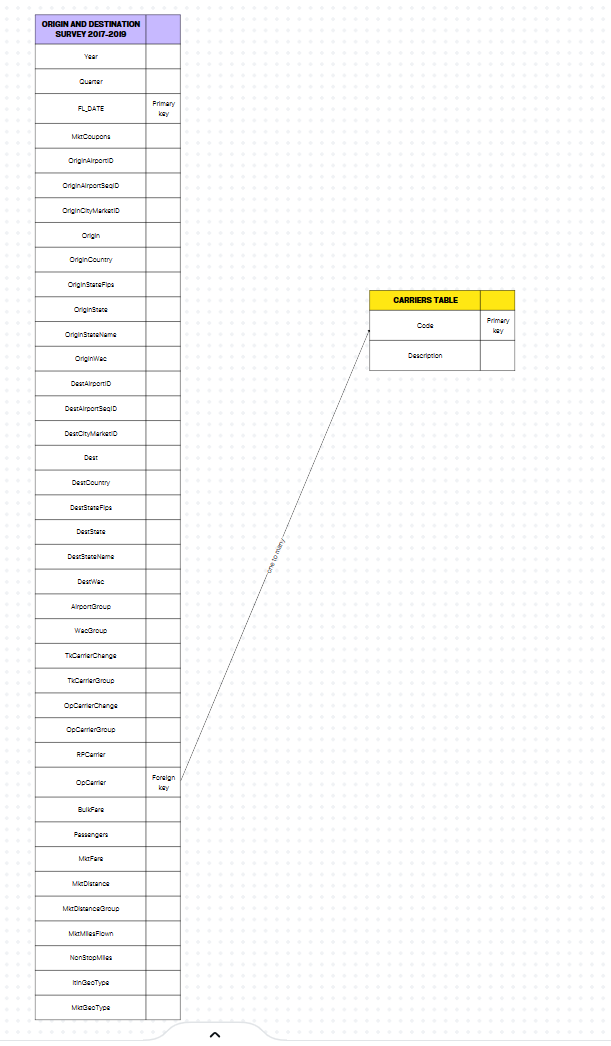
Data can also be used to study trends and patterns in the travel industry, which can act as a valuable resource for researchers and analysts.

The US holidays dataset contains information about when and what holidays are occurring from 2004 to 2021. Knowing when the holidays are, may lead to discovery of new insights between holidays and ticket fare, holidays and state, etc.

Data Catalogue

*(Describe the structures and fields of your datasets/ data sources. Draw ER diagrams to show the relationship between the various datasets if the datasets that you are using is related to other datasets. For example, if dataset A contains column B whose meaning can only be understood by interpreting column Z in the dataset)*

<https://www.canva.com/design/DAFoCCEs3Uc/SQhd1Oa_ks6O6Ggv3BGEPw/edit?utm_content=DAFoCCEs3Uc&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton>



Miscellaneous

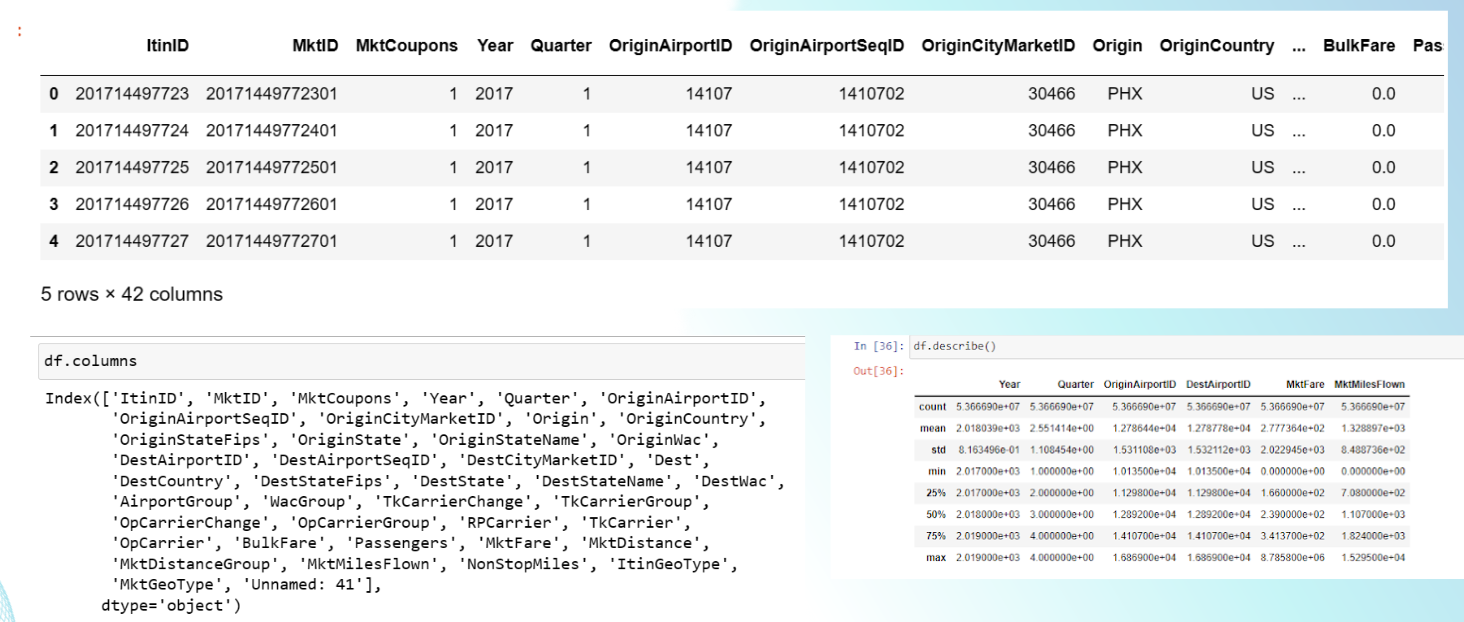
*(anything else you wish to add about your project can be included here)*

Data dictionary

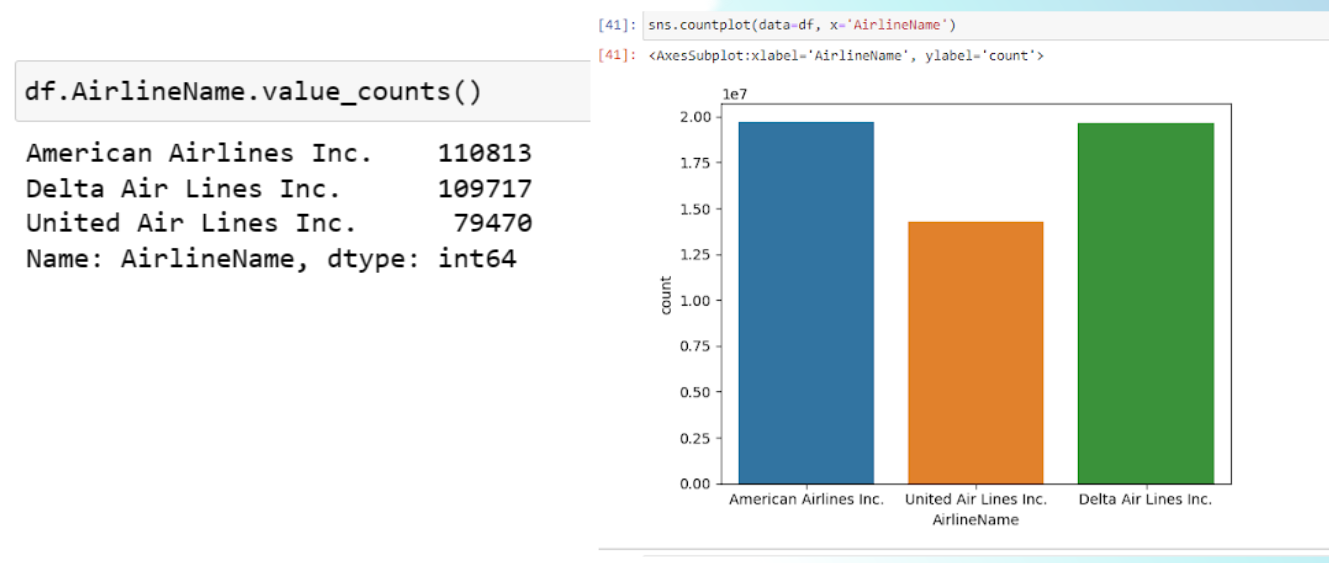
| **Field Name** | **Description** |
| --- | --- |
| ItinID | Itinerary ID |
| MktID | Market ID |
| [MktCoupons](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=a7zor4%FD1s%FDP172105%FDv0%FD6ur%FDZn4xr6&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_QOEO_Pbhcbaf&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=ZNeXRg_Pbhcbaf) | Number of Coupons in the Market |
| Year | Year |
| [Quarter](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=d7n46r4%FD%FLE-H%FM&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_dhNegRef&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=dhNegRe) | Quarter (1-4) |
| [OriginAirportID](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b4vtv0%FDNv42146%FP%FDNv42146%FDVQ.%FDN0%FDvqr06vsvpn6v10%FD07zor4%FDn55vt0rq%FDoB%FDhf%FDQbg%FD61%FDvqr06vsB%FDn%FD70v37r%FDnv42146.%FD%FDh5r%FD6uv5%FDsvryq%FDs14%FDnv42146%FDn0nyB5v5%FDnp4155%FDn%FD4n0tr%FD1s%FDBrn45%FDorpn75r%FDn0%FDnv42146%FDpn0%FDpun0tr%FDv65%FDnv42146%FDp1qr%FDn0q%FDnv42146%FDp1qr5%FDpn0%FDor%FD4r75rq.&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_NVecbeg_VQ&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=beVTVa_NVecbeg_VQ) | Origin Airport, Airport ID. An identification number assigned by US DOT to identify a unique airport. Use this field for airport analysis across a range of years because an airport can change its airport code and airport codes can be reused. |
| [OriginAirportSeqID](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b4vtv0%FDNv42146%FP%FDNv42146%FDfr37r0pr%FDVQ.%FDN0%FDvqr06vsvpn6v10%FD07zor4%FDn55vt0rq%FDoB%FDhf%FDQbg%FD61%FDvqr06vsB%FDn%FD70v37r%FDnv42146%FDn6%FDn%FDtv8r0%FD21v06%FD1s%FD6vzr.%FD%FDNv42146%FDn664vo76r5%FP%FD57pu%FDn5%FDnv42146%FD0nzr%FD14%FDp114qv0n6r5%FP%FDznB%FDpun0tr%FD18r4%FD6vzr.&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_NVecbeg_fRd_VQ&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=beVTVa_NVecbeg_fRd_VQ) | Origin Airport, Airport Sequence ID. An identification number assigned by US DOT to identify a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time. |
| [OriginCityMarketID](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b4vtv0%FDNv42146%FP%FDPv6B%FDZn4xr6%FDVQ.%FDPv6B%FDZn4xr6%FDVQ%FDv5%FDn0%FDvqr06vsvpn6v10%FD07zor4%FDn55vt0rq%FDoB%FDhf%FDQbg%FD61%FDvqr06vsB%FDn%FDpv6B%FDzn4xr6.%FD%FDh5r%FD6uv5%FDsvryq%FD61%FDp1051yvqn6r%FDnv421465%FD5r48v0t%FD6ur%FD5nzr%FDpv6B%FDzn4xr6.&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_PVgl_ZNeXRg_VQ&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=beVTVa_PVgl_ZNeXRg_VQ) | Origin Airport, City Market ID. City Market ID is an identification number assigned by US DOT to identify a city market. Use this field to consolidate airports serving the same city market. |
| [Origin](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b4vtv0%FDNv42146%FDP1qr&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_NVecbeg&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=beVTVa) | Origin Airport Code |
| OriginCountry | Origin Airport, Country Code |
| [OriginStateFips](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b4vtv0%FDNv42146%FP%FDf6n6r%FDSVcf%FDP1qr&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_fgNgR_SVcf&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=beVTVa_fgNgR_SVcf) | Origin Airport, State FIPS Code |
| [OriginState](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b4vtv0%FDNv42146%FP%FDf6n6r%FDP1qr&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_fgNgR_NOe_NiVNgVba&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=beVTVa_fgNgR_NOe) | Origin Airport, State Code |
| OriginStateName | Origin State Name |
| [OriginWac](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b4vtv0%FDNv42146%FP%FDj14yq%FDN4rn%FDP1qr&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_jbeYQ_NeRN_PbQRf&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=beVTVa_jNP) | Origin Airport, World Area Code |
| [DestAirportID](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qr56v0n6v10%FDNv42146%FP%FDNv42146%FDVQ.%FDN0%FDvqr06vsvpn6v10%FD07zor4%FDn55vt0rq%FDoB%FDhf%FDQbg%FD61%FDvqr06vsB%FDn%FD70v37r%FDnv42146.%FD%FDh5r%FD6uv5%FDsvryq%FDs14%FDnv42146%FDn0nyB5v5%FDnp4155%FDn%FD4n0tr%FD1s%FDBrn45%FDorpn75r%FDn0%FDnv42146%FDpn0%FDpun0tr%FDv65%FDnv42146%FDp1qr%FDn0q%FDnv42146%FDp1qr5%FDpn0%FDor%FD4r75rq.&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_NVecbeg_VQ&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QRfg_NVecbeg_VQ) | Destination Airport, Airport ID. An identification number assigned by US DOT to identify a unique airport. Use this field for airport analysis across a range of years because an airport can change its airport code and airport codes can be reused. |
| [DestAirportSeqID](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qr56v0n6v10%FDNv42146%FP%FDNv42146%FDfr37r0pr%FDVQ.%FDN0%FDvqr06vsvpn6v10%FD07zor4%FDn55vt0rq%FDoB%FDhf%FDQbg%FD61%FDvqr06vsB%FDn%FD70v37r%FDnv42146%FDn6%FDn%FDtv8r0%FD21v06%FD1s%FD6vzr.%FD%FDNv42146%FDn664vo76r5%FP%FD57pu%FDn5%FDnv42146%FD0nzr%FD14%FDp114qv0n6r5%FP%FDznB%FDpun0tr%FD18r4%FD6vzr.&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_NVecbeg_fRd_VQ&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QRfg_NVecbeg_fRd_VQ) | Destination Airport, Airport Sequence ID. An identification number assigned by US DOT to identify a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time. |
| [DestCityMarketID](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qr56v0n6v10%FDNv42146%FP%FDPv6B%FDZn4xr6%FDVQ.%FDPv6B%FDZn4xr6%FDVQ%FDv5%FDn0%FDvqr06vsvpn6v10%FD07zor4%FDn55vt0rq%FDoB%FDhf%FDQbg%FD61%FDvqr06vsB%FDn%FDpv6B%FDzn4xr6.%FD%FDh5r%FD6uv5%FDsvryq%FD61%FDp1051yvqn6r%FDnv421465%FD5r48v0t%FD6ur%FD5nzr%FDpv6B%FDzn4xr6.&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_PVgl_ZNeXRg_VQ&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QRfg_PVgl_ZNeXRg_VQ) | Destination Airport, City Market ID. City Market ID is an identification number assigned by US DOT to identify a city market. Use this field to consolidate airports serving the same city market. |
| [Dest](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qr56v0n6v10%FDNv42146%FDP1qr&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_NVecbeg&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QRfg) | Destination Airport Code |
| DestCountry | Destination Airport, Country Code |

| [DestStateFips](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qr56v0n6v10%FDNv42146%FP%FDf6n6r%FDSVcf%FDP1qr&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_fgNgR_SVcf&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QRfg_fgNgR_SVcf) | Destination Airport, State FIPS Code |
| --- | --- |
| [DestState](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qr56v0n6v10%FDNv42146%FP%FDf6n6r%FDP1qr&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_fgNgR_NOe_NiVNgVba&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QRfg_fgNgR_NOe) | Destination Airport, State Code |
| DestStateName | Destination State Name |
| [DestWac](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qr56v0n6v10%FDNv42146%FP%FDj14yq%FDN4rn%FDP1qr&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_jbeYQ_NeRN_PbQRf&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QRfg_jNP) | Destination Airport, World Area Code |
| AirportGroup | Airport Group |
| WacGroup | World Area Code Group |
| [TkCarrierChange](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=gvpxr6v0t%FDPn44vr4%FDPun0tr%FDV0qvpn614%FD%FLE%GQlr5%FM&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_lRfab_eRfc&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=gX_PNeeVRe_PUNaTR) | Ticketing Carrier Change Indicator (1=Yes) |
| TkCarrierGroup | Ticketing Carrier Group |
| [OpCarrierChange](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b2r4n6v0t%FDPn44vr4%FDPun0tr%FDV0qvpn614%FD%FLE%GQlr5%FM&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_lRfab_eRfc&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=bc_PNeeVRe_PUNaTR) | Operating Carrier Change Indicator (1=Yes) |
| OpCarrierGroup | Operating Carrier Group |
| [RPCarrier](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=er2146v0t%FDPn44vr4%FDP1qr&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_PNeeVRef&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=eRcbegVaT_PNeeVRe) | Reporting Carrier Code |
| [TkCarrier](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=gvpxr6v0t%FDPn44vr4%FDP1qr%FDs14%FDb0-yv0r%FDV6v0r4n4vr5%FD%FL16ur49v5r%FDr37ny%FD61%FDMM%FM&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_PNeeVRef&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=gVPXRg_PNeeVRe) | Ticketing Carrier Code for On-line Itineraries (otherwise equal to 99) |
| [OpCarrier](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=b2r4n6v0t%FDPn44vr4%FDP1qr%FDs14%FDb0-yv0r%FDV6v0r4n4vr5%FD%FL16ur49v5r%FDr37ny5%FD61%FDMM%FM&Svryq_gB2r=Pun4&Y11x72_gnoyr=Y_PNeeVRef&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=bcReNgVaT_PNeeVRe) | Operating Carrier Code for On-line Itineraries (otherwise equals to 99) |
| [BulkFare](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=O7yx%FDSn4r%FDV0qvpn614%FD%FLE%GQlr5%FM&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_QOEO_OhYXSNeR&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=OhYX_SNeR) | Bulk Fare Indicator (1=Yes) |
| Passengers | Number of Remaining Tickets |
| MktFare | Market Fare (ItinYield\*MktMilesFlown) |
| MktDistance | Market Distance (Including Ground Transport) |
| [MktDistanceGroup](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Qv56n0pr%FDT4172%FP%FDv0%FDIDD%FDZvyr%FDV06r48ny5&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_QVfgNaPR_Tebhc_IDD&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=QVfgNaPR_Tebhc) | Distance Group, in 500 Mile Intervals |
| MktMilesFlown | Market Miles Flown (Track Miles) |
| NonStopMiles | Non-Stop Market Miles (Using Radian Measure) |
| [ItinGeoType](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=V6v0r4n4B%FDTr1t4n2uB%FDgB2r&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_TRbglcR&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=VgVa_TRb_glcR) | Itinerary Geography Type |
| [MktGeoType](https://www.transtats.bts.gov/FieldInfo.asp?Svryq_Qr5p=Zn4xr6%FDTr1t4n2uB%FDgB2r&Svryq_gB2r=a7z&Y11x72_gnoyr=Y_TRbglcR&gnoyr_VQ=FHK&flf_gnoyr_anzr=g_QOEO_ZNeXRg&fB5_Svryq_anzr=ZXg_TRb_glcR) | Market Geography Type |

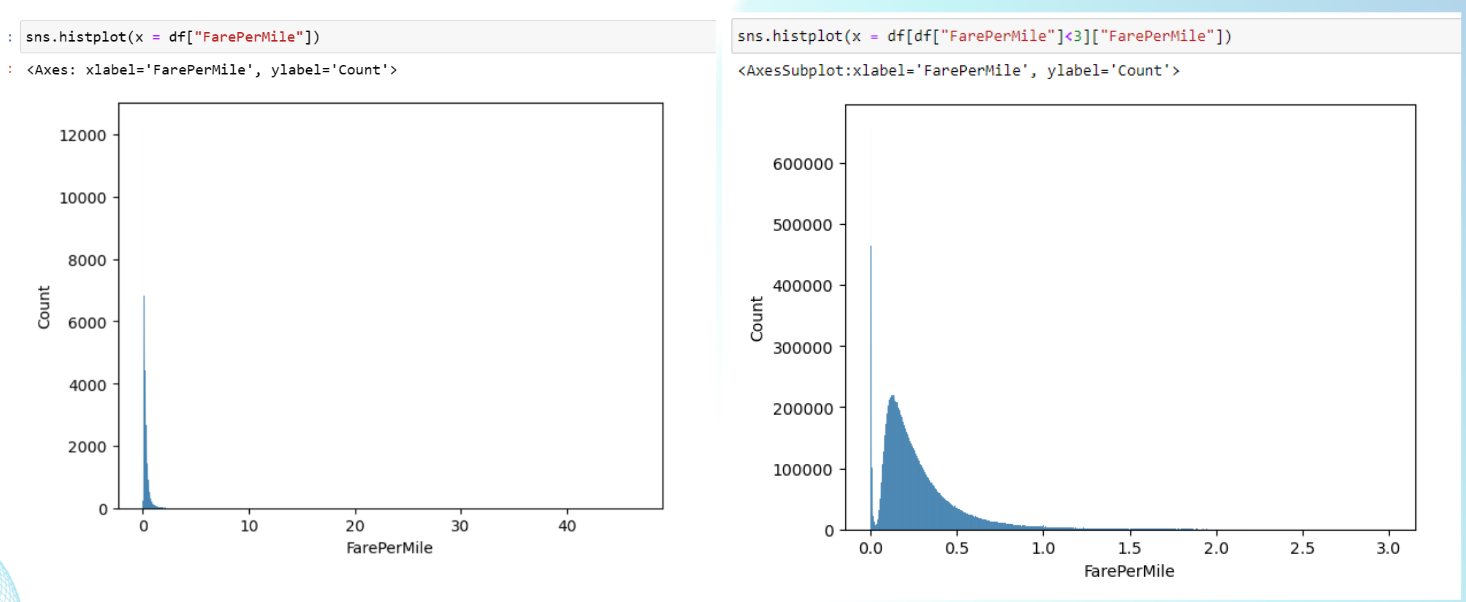
Exploratory Data Analysis



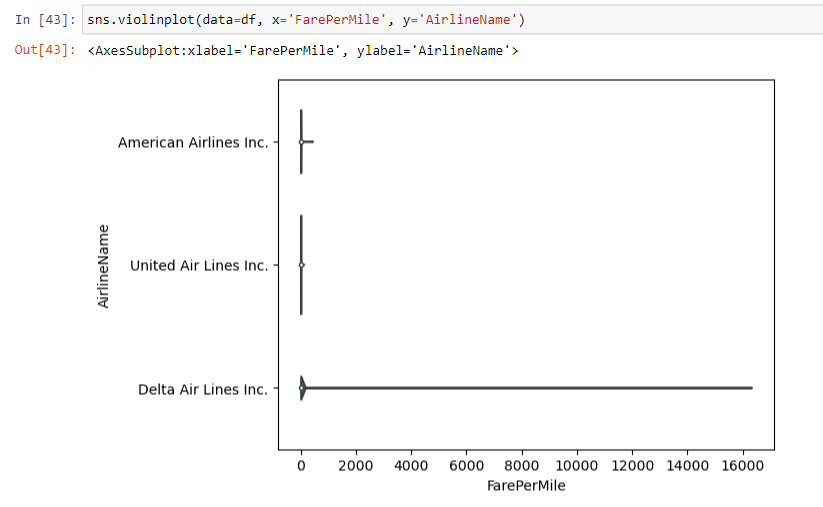
A glimpse of how the data looks like, columns available and aggregated statistics of numerical columns.



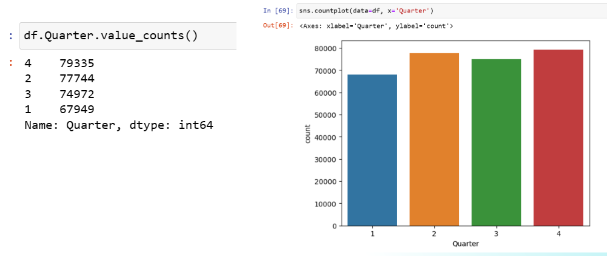
American airlines has the highest number of records, followed by Delta airlines and United airlines



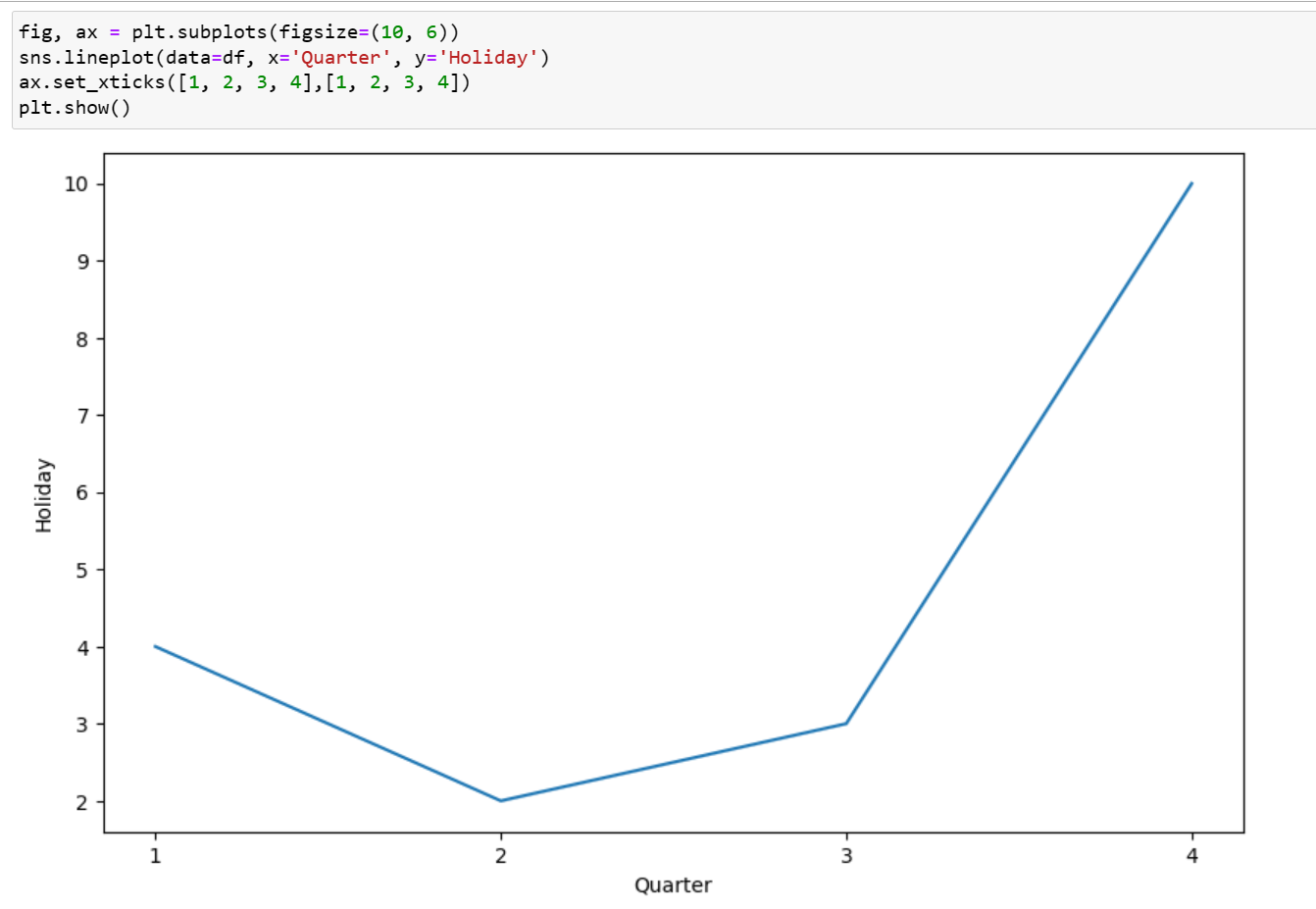
Distribution of fare per mile before and after removing outliers. It is evident that there are a large number of outliers in fare per mile, indicating the presence of anomalies.



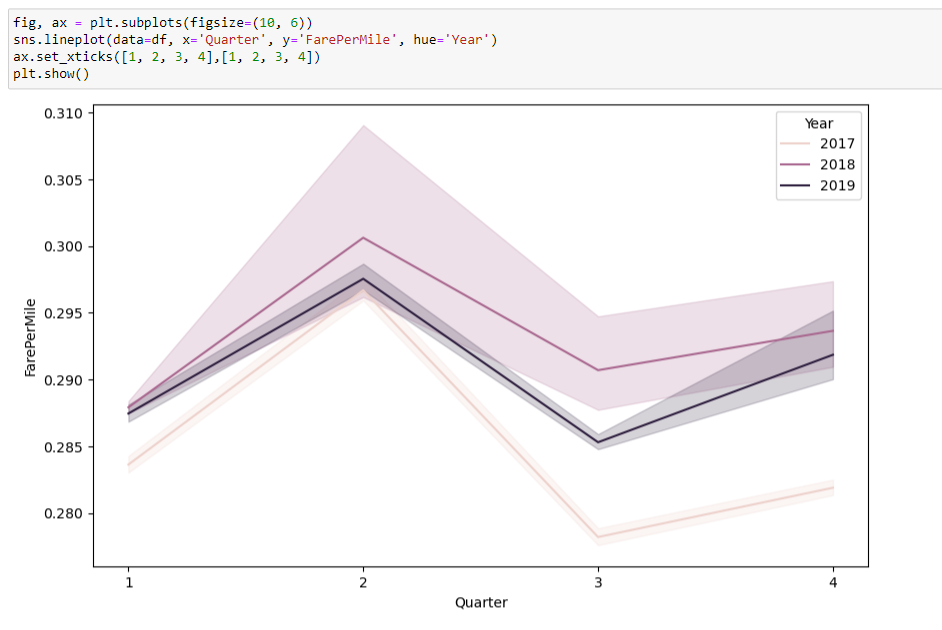
Delta airlines has the largest spread in fare per mile, indicating that Delta airlines is responsible for the previously seen outliers in fare per mile.



Quarter 4 has the highest sales, followed by 2, 3, and 1.

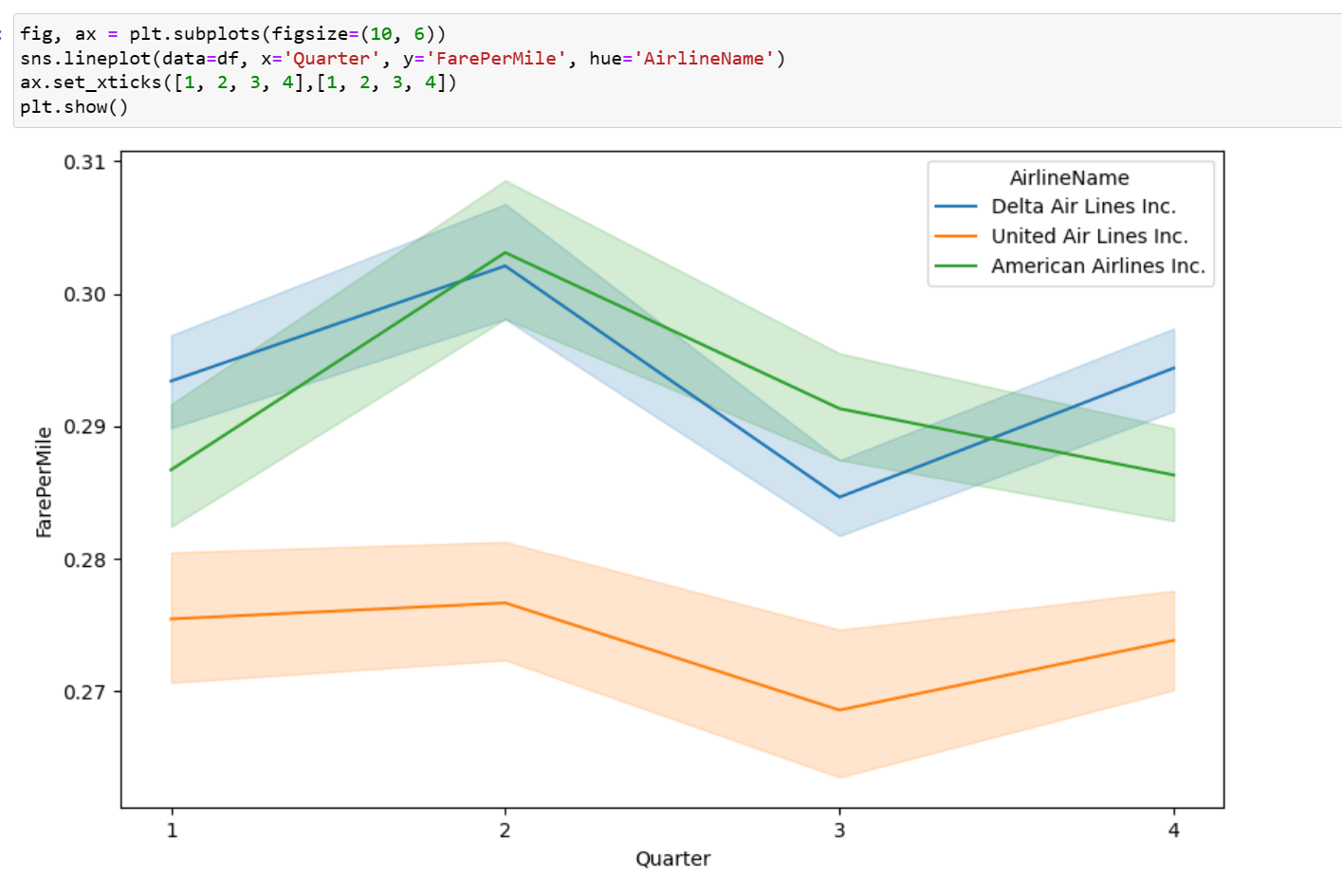


Quarter 4 has the highest number of holidays, followed by 1, 3 and 2



2017 has the biggest fluctuations, while 2018 and 2019 showed similar difference in quarters

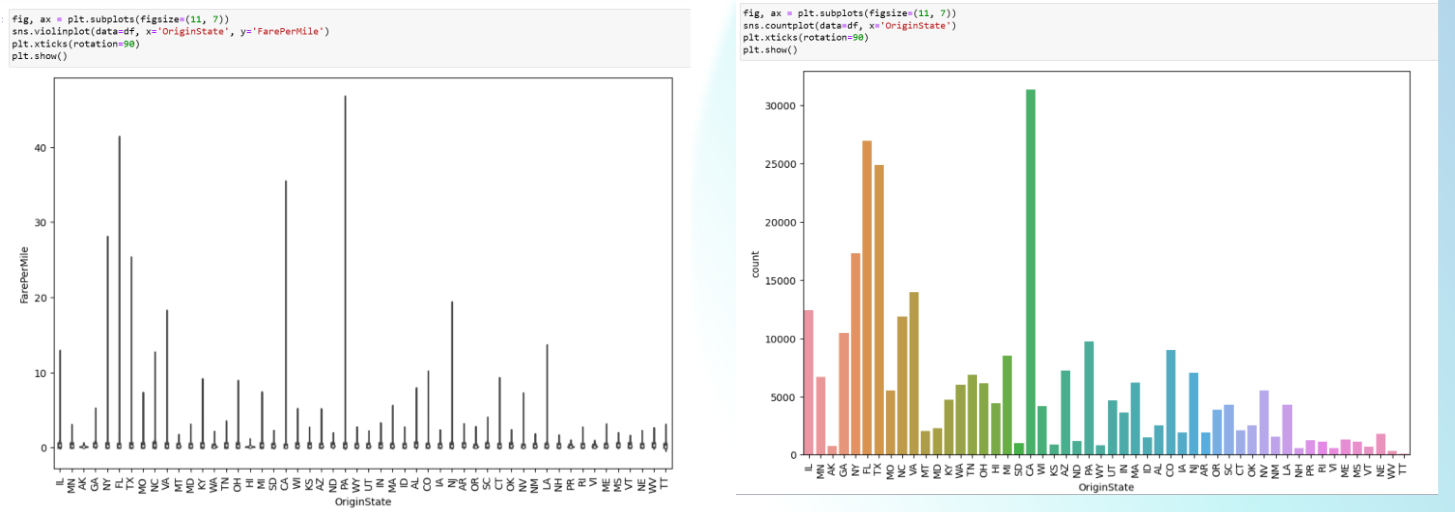
Overall, quarter 2 has the highest fare per mile. This could be caused by high demand for tickets but low number of public holidays as seen in the previous charts.



American airlines follow the general trend, where quarter 2 has the highest fare per mile.

However, even though United airlines has the highest fare per mile in quarter 2 as well, they prefer to maintain a low fare per mile throughout the whole year.

In contrast, Delta airlines has the highest fare per mile in quarter 4, 1 and 2, but has a significant drop in quarter 3 where they price match to the other airlines.



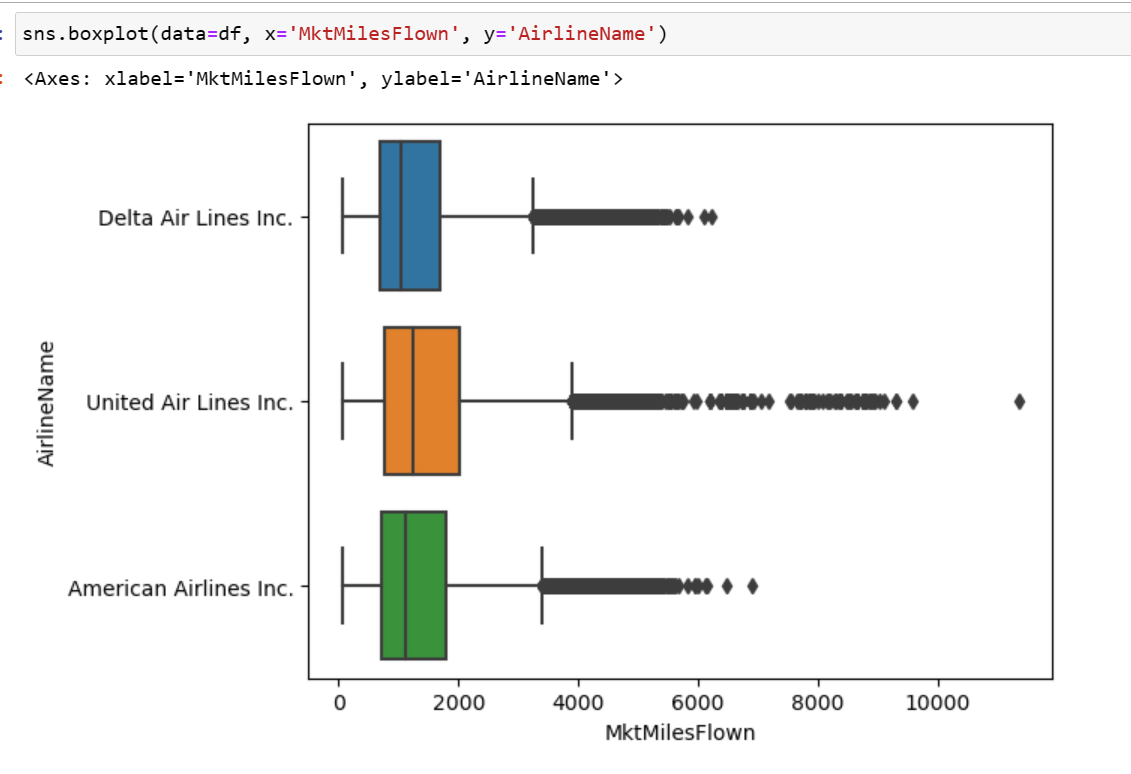
PA has the widest spread of fare per mile, but CA has the highest count of tickets sold.

Whereas, FL has the second highest count and second widest spread of fare per mile.

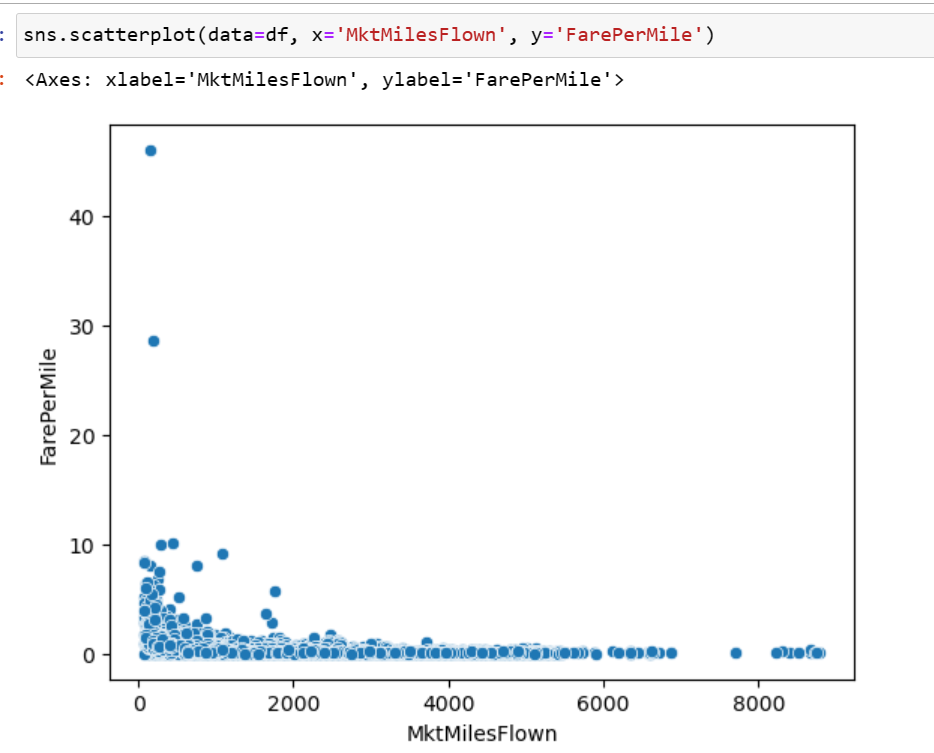
MI has the widest spread of fare per mile, but CA has the highest count of tickets sold.

Whereas, FL has the second highest count and second widest spread of fare per mile.

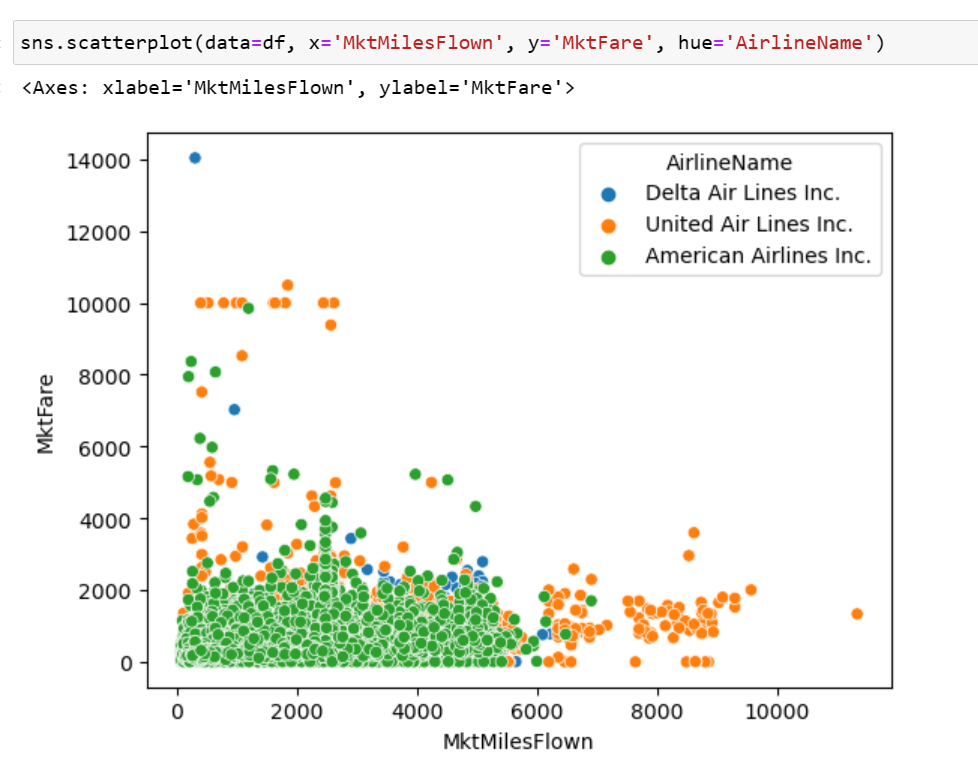
The insights for Fare per mile by destination state is similar to those of origin state.



Even though United Airlines has the lowest fare per miles and the lowest number of tickets sold, it has overall more miles flown than Delta or American airlines

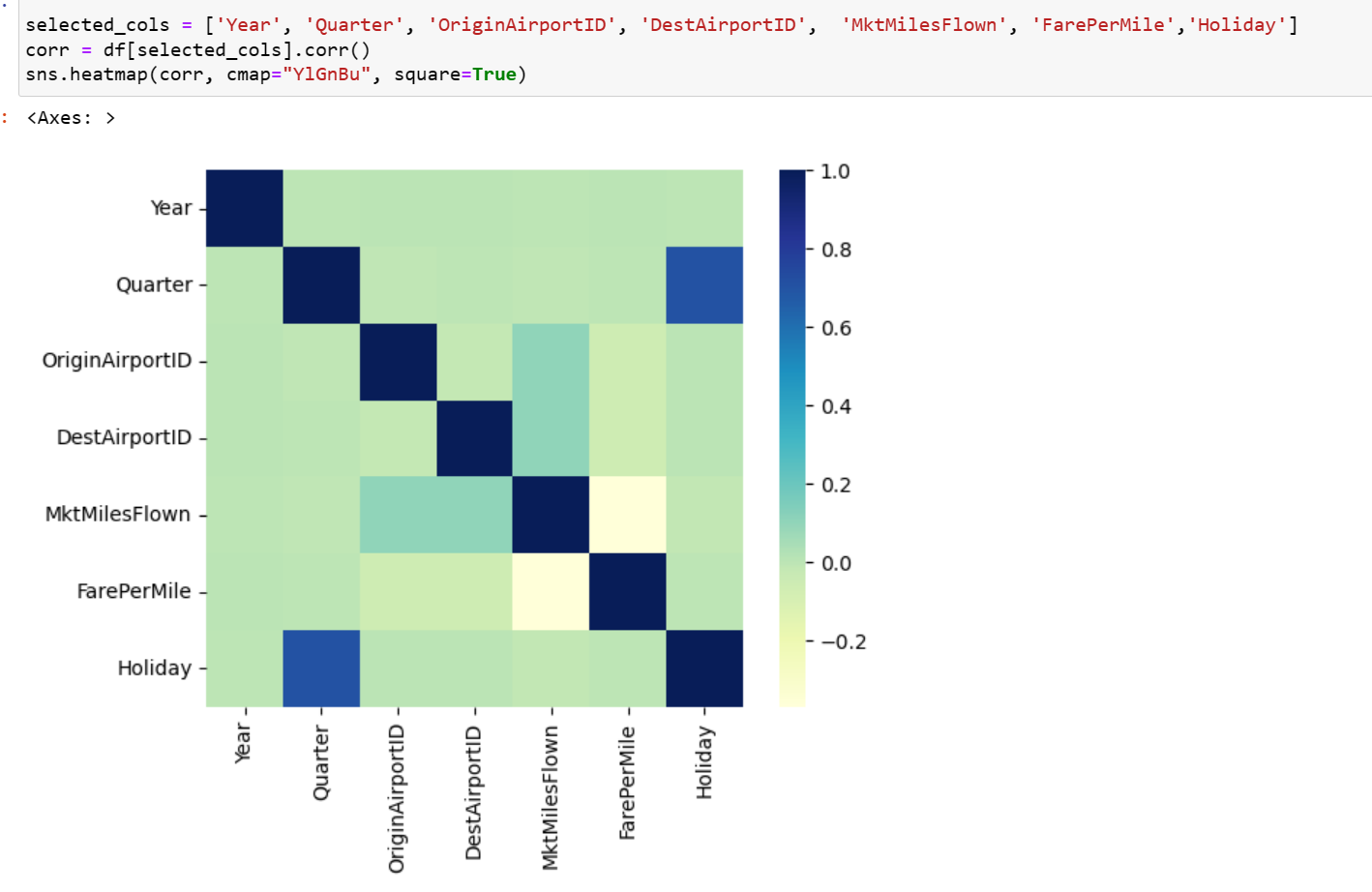


Miles flown does not have a strong correlation with fare per mile, and it is evident that there are a large number of outliers in fare per mile that have low miles flown.



Miles flown does not have a strong correlation with market fare, and it is evident that there are a large number of outliers in market fare that have low miles flown.

United has most number of outliers in both miles flown and market fare



Fare Per mile has stronger correlation with Year, Quarter and Holiday, whereas Market miles flown has the weakest correlation.