

# project

June 17, 2025

## 1 Will Hawaii Sink or Swim without Tourists?

### 1.1 Summary

Being born and raised in the Hawaiian Islands, I've seen firsthand how tourism can feel like a double-edged sword. While it brings in revenue, it also strains local resources, inflates housing costs, and in crisis moments like the COVID-19 shutdowns, exposes how fragile Hawaii's economy and general well-being can be. Driven by this view, this project investigates exactly how dependent Hawaii's economy is on tourism and whether the state could potentially have a more sustainable future without tourism involved. Drawing on federal and state economic data, it analyzes shifts in GDP, employment, and visitor trends in regards to the pandemic. By quantifying tourism's true impact, the project aims to discover Hawaii's true dependence on tourism as a whole, safeguard Native Hawaiian culture, and protect our āina for generations to come.

---

### 1.2 Research Questions

1. **How significantly does Hawaii's economy rely on tourism-related industries?**
  - Hawaii's economy does infact rely heavily on tourism. It could be seen that Hawaii has consistently made up about 8-12% of Hawaii's economy from 2005 to 2024, which is huge margin of the state's GDP. Not only that, but between all the other industries, these two industries combined make up the third most of Hawaii's total GDP.
2. **What economic trends/declines did Hawaii experience during the COVID-19 pandemic when tourism was severely restricted?**
  - As a result of visitor arrivals dropping during the pandemic in 2020 Q2, the number of civilians had dropped, as well as the GDP from tourism-related industries.
3. **How did the decline in visitor arrivals during the COVID-19 pandemic affect state-level housing affordability?**
  - Visitor arrivals may have some sort of association/correlation with the increase in the cost of housing. As the number of visitors rise, so does the median housing price. This may likely be from broader economic growth. However, this correlation does not necessarily prove causation, as the median of the housing price continues to increase as the number of visitors increases. Thus, tourism is not a direct cause of housing price inflation. This housing price inflation could be caused by other demographic or economic factors.

## 1.3 Challenge Goals

### 1.3.1 1. Multiple Datasets

My project will incorporate **three distinct datasets** to analyze Hawai‘i’s economic dependence on tourism and the potential of alternative industries:

- **DBEDT (Hawai‘i Department of Business, Economic Development & Tourism)** – Three distinct datasets from this website:
  - Statewide and county-specific data on employment/unemployment rates.
  - Data on tourism arrivals, counts, spending, etc.
  - Housing median prices in certain counties of Hawaii.
- **UHERO (University of Hawai‘i Economic Research Organization)** – Offers overall and industry-specific GDP data , helping me assess the impact of both tourism and non-tourism industries.

By combining these datasets, I can trace links between visitor trends, GDP, and employment especially before, during, and after the COVID-19 pandemic.

---

### 1.3.2 2. New Library

I will use the Python library **Plotly**, which was not covered in our coursework, to build interactive visualizations (i.e. dynamic line graphs and hoverable bar charts). Plotly will let viewers:

- Zoom in on timelines
- Hover to see exact values
- Compare tourism with emerging industries at a glance

Although I have experience with other libraries like `matplotlib`, this will be my first time using Plotly. Learning its interactive features will make economic trends clearer for both general audiences and policy stakeholders.

---

By pairing these new technical skills with multi-source economic data, I aim to answer not just whether Hawai‘i relies on tourism, but whether it must. The results could inform broader discussions about sustainability, economic resilience, and cultural preservation. “

## 1.4 Collaboration and Conduct

Students are expected to follow Washington state law on the [Student Conduct Code for the University of Washington](#). In this course, students must:

- Indicate on your submission any assistance received, including materials distributed in this course.
- Not receive, generate, or otherwise acquire any substantial portion or walkthrough to an assessment.
- Not aid, assist, attempt, or tolerate prohibited academic conduct in others.

Update the following code cell to include your name and list your sources. If you used any kind of computer technology to help prepare your assessment submission, include the queries and/or prompts. Submitted work that is not consistent with sources may be subject to the student conduct process.

```
[1]: your_name = "Dylan Domingo"
sources = [
    "https://dbedt.hawaii.gov/visitor/tourismdata/", # tourism dataset

    "https://dbedt.hawaii.gov/economic/datawarehouse/", # employment/
    ↪unemployment, housing prices datasets

    "https://data.uhero.hawaii.edu/#/category?id=21&data_list_id=20", # gdp
    ↪dataset

    "https://courses.cs.washington.edu/courses/cse163/25sp/2025/04/14/
    ↪data-frames/", # DataFrames lecture, used to review

    ↪ # methods to manipulate

    "https://plotly.com/python/", # plotly library, used to learn how plotly
    ↪works and documentation

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.astype.
    ↪html", # how to cast columns to certain object types

    "https://numpy.org/doc/2.2/reference/arrays.dtypes.html", # how dtypes work
    ↪and the significance for object checking in DataFrames

    "https://docs.python.org/3/tutorial/errors.html", # how exception handling
    ↪works in python

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.drop.html",
    ↪# how to drop columns in pandas, used to drop unnecessary

    ↪# columns to data analysis

    "https://www.w3schools.com/python/ref_string_replace.asp", # how to replace
    ↪substrings in python, used in my clean

    ↪# method to

    ↪standardize values

    "https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html", # how
    ↪read_csv works, found out about thousands

    ↪#

    ↪parameter to delimit commas
```

```

    "https://www.geeksforgeeks.org/stringio-module-in-python/", # how StringIO
↳works, which helped with testing my functions

    "https://pandas.pydata.org/docs/reference/api/pandas.melt.html", # how melt
↳works, particularly to melt wide to long forms for Plotly graphs

    "https://courses.cs.washington.edu/courses/cse163/25sp/2025/04/16/
↳groupby-and-indexing/#Groupby-in-Pandas", # groupby lecture to

↳                                     # understand when i should use
↳                                     # groupby

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sum.html", #
↳how .sum() works to sum rows/columns of a DataFrame

    "https://courses.cs.washington.edu/courses/cse163/25sp/2025/04/25/objects/
↳#Type-annotations", # how lambda functions work, used within

↳                                     # apply functions to DataFrames

    "https://courses.cs.washington.edu/courses/cse163/25sp/2025/05/14/
↳dissolve-intersect-and-join/#Dissolve,-Intersect,-and-Join", # lecture on

↳                                     # how merge works
↳                                     # used in cases
↳                                     # where there
↳                                     # were multiple
↳                                     # datasets

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rename.
↳html", # how to rename columns in pandas

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.replace.
↳html#pandas.DataFrame.replace", # how replace works in pandas

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.merge.
↳html#pandas.DataFrame.merge", # review on how merge works, used in cases

↳                                     # where there were multiple datasets

```

```

    "https://plotly.com/python/reference/layout/yaxis/", # how to adjust a
    ↪y-axis on plotly

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.values.
    ↪html", # how to access all values of a DataFrame

    "https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.columns.
    ↪html#pandas.DataFrame.columns", # how to access all columns of

    ↪                                # a DataFrame

    "https://usafacts.org/articles/
    ↪which-states-have-the-highest-and-lowest-rates-of-homelessness/", #
    ↪assessing hawaii's homelessness rate

    "https://www.hawaii-guide.com/hawaii-tourism-statistics", # assessing
    ↪hawaii's most visited island, Oahu
]

assert your_name != "", "your_name cannot be empty"
assert ... not in sources, "sources should not include the placeholder ellipsis"
assert len(sources) >= 6, "must include at least 6 sources, inclusive of
    ↪lectures and sections"

```

## 1.5 Data Setting and Methods

To begin, I imported all the necessary libraries including `pandas` for data manipulation and `plotly` for interactive visualization.

```

[2]: !pip install pandas
    !pip install plotly
    !pip install numpy

```

```

Requirement already satisfied: pandas in /opt/conda/lib/python3.11/site-packages
(2.2.2)
Requirement already satisfied: numpy>=1.23.2 in /opt/conda/lib/python3.11/site-
packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/opt/conda/lib/python3.11/site-packages (from pandas) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.11/site-
packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.11/site-
packages (from pandas) (2024.1)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.11/site-
packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

```

[notice] A new release of pip is

```

available: 25.0.1 -> 25.1.1
[notice] To update, run:
pip install --upgrade pip
Collecting plotly
  Using cached plotly-6.1.2-py3-none-any.whl.metadata (6.9 kB)
Collecting narwhals>=1.15.1 (from plotly)
  Downloading narwhals-1.43.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: packaging in /opt/conda/lib/python3.11/site-
packages (from plotly) (24.2)
Using cached plotly-6.1.2-py3-none-any.whl (16.3 MB)
Downloading narwhals-1.43.0-py3-none-any.whl (362 kB)
Installing collected packages: narwhals, plotly
Successfully installed narwhals-1.43.0 plotly-6.1.2

[notice] A new release of pip is
available: 25.0.1 -> 25.1.1
[notice] To update, run:
pip install --upgrade pip
Requirement already satisfied: numpy in /opt/conda/lib/python3.11/site-packages
(1.26.4)

[notice] A new release of pip is
available: 25.0.1 -> 25.1.1
[notice] To update, run:
pip install --upgrade pip

```

```

[3]: import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
import numpy as np
from io import StringIO

```

I then created a function to help clean and reformat the data. Here are the functions:

```

[4]: def load_and_clean_csv(filepath):
    """
    Loads CSV data from a file, standardizes column names to lowercase with
    ↪underscores,
    removes commas from string numbers, and converts columns to float where
    ↪possible.
    Returns the cleaned DataFrame.
    """
    df = pd.read_csv(filepath, thousands=",")
    df.columns = df.columns.str.lower().str.strip().str.replace(" ", "_").str.
    ↪replace("\\"", "")

    for col in df.columns:

```

```

    if df[col].dtype == "object":
        try:
            df[col] = df[col].astype(float)
        except ValueError:
            continue
    return df

```

I then loaded in datasets from the DBEDT (tourism and employment stats), and UHERO (total and industry-specific GDP data) as CSV files. I made sure to set indexes, drop unnecessary columns (like units of measure) as well for easier indexing.

```

[5]: # load csv files
employment_data = load_and_clean_csv("employment.csv").set_index(["indicator", "area"])
employment_data = employment_data.drop("units", axis=1)

tourism_data = load_and_clean_csv("tourism.csv").set_index("indicator")
tourism_data = tourism_data.drop(["market", "units"], axis=1)

gdp_data = load_and_clean_csv("gdp.csv").set_index("series")

housing_data = load_and_clean_csv("housing_prices.csv").set_index("indicator")
housing_data = housing_data.drop("units", axis=1)

# display(employment_data)
# display(tourism_data)
# display(gdp_data)
# display(housing_data)

```

## 1.6 Results

### 1.6.1 How significantly does Hawaii's economy rely on tourism-related industries?

To develop an answer to this question, I first identified tourism-related industries within my dataset. I came to a conclusion regarding which industries were related to tourism: Arts, Entertainment, and Recreation & Accommodation and Food Services. - I chose the Arts, Entertainment, and Recreation industry to be included within my analysis as this most likely has to deal with traditional performances in Hawaii, This is including but not limited to hula dancing, luaus, or maybe even museums containing traditional and ancient Hawaiian art. Tourists are most likely to visit such attractions, which makes up most of the GDP within this industry. - I chose the Accommodation and Food Services industry as this industry has to deal with lodging and hotel services. Most times, tourists stay within hotels during their stay, so the tourists make up most of the GDP of the Accommodation and Food Services industry.

Next, I needed some way to quantify tourist-related industry impact on Hawaii's economy. I decided on finding out the total share that these tourism-related industries hold of the state's total GDP. This is called the Tourism Share, or the percentage of the total GDP that tourism industries make up:

$$\begin{aligned} \text{Let } GDP_T &= \text{total GDP} \\ \text{Let } GDP_1 &= \text{GDP of Arts, Entertainment, and Recreation} \\ \text{Let } GDP_2 &= \text{GDP of Accommodation and Food Services} \\ \text{Tourism Share} &= \frac{GDP_1 + GDP_2}{GDP_T} \cdot (100) \end{aligned}$$

I found the tourism share across all quarters spanning from 2005 Q1 to 2024 Q4, then plotted it on a timeline plot to assess tourism's affect on the economy of Hawaii.

```
[6]: def compute_industry_share(whole_gdp, industries_gdp):
    """
    Takes in a Series of a state's whole GDP and a list of Series of
    industries' sole GDP stats,
    and returns a percentage of how much the given industries make up of the
    total GDP.
    """
    industry_sum = 0
    for industry in industries_gdp:
        industry_sum += industry
    return (industry_sum / whole_gdp) * 100

total_gdp = gdp_data.loc[" Total Gross Domestic Product ($Mil)"]
arts_ent = gdp_data.loc[" GDP: Arts, Entertainment, and Recreation ($Mil)"]
accom_food = gdp_data.loc[" GDP: Accommodation and Food Services ($Mil)"]

tourism_share = compute_industry_share(total_gdp, [arts_ent, accom_food]) #
    compute tourism_share

plot_df = pd.DataFrame({ # create a separate DataFrame for plotting
    "Quarter": gdp_data.columns,
    "Tourism Share (%)": tourism_share.values,
}).dropna()

fig_share = px.line( # plot gdp share rate
    plot_df,
    x="Quarter",
    y="Tourism Share (%)",
    title="Tourism-Related Industries as % of Hawaii's Total GDP",
    markers=True,
)
fig_share.update_layout(
    xaxis_title="Quarter",
    yaxis_title="Percent of Total GDP"
)
fig_share.show(renderer="notebook")
```



```

## testing compute_industry_share ##
mock_gdp = StringIO(
    '''
    "series",2023_q1,2023_q2
    "total_gdp",1000,1100
    "industry_one_gdp",120,130
    "industry_two_gdp",180,190
    '''
)

mock_gdp_data = pd.read_csv(mock_gdp).set_index("series")
mock_total_gdp = mock_gdp_data.loc["total_gdp"]
mock_industry_one_gdp = mock_gdp_data.loc["industry_one_gdp"]
mock_industry_two_gdp = mock_gdp_data.loc["industry_two_gdp"]
mock_tourism_share = compute_industry_share(
    mock_total_gdp,
    [mock_industry_one_gdp, mock_industry_two_gdp]
)

expected_q1 = ((120 + 180) / 1000) * 100 # 30.0
expected_q2 = ((130 + 190) / 1100) * 100 # 29.0909

assert abs(mock_tourism_share["2023_q1"] - expected_q1) < 1e-6, "Q1 share_
↳incorrect"
assert abs(mock_tourism_share["2023_q2"] - expected_q2) < 1e-6, "Q2 share_
↳incorrect"

```

Upon plotting my findings, I found that the Tourism Share stayed at a consistent 8-12% of the time. I then grew curious about how that compares to other industries. so I created a multi line plot of all other industries GDPs and how it compares to the tourism industry.

```

[7]: dropped_gdp_data = gdp_data.drop([ # first drop total and real gdp
    " Total Gross Domestic Product  ($Mil)",
    " Real Gross Domestic Product  (2024$Mil)",
])

df_long = ( # melt DataFrame from wide into long for easier plotting
    dropped_gdp_data
    .reset_index()
    .melt(id_vars="series", var_name="quarter", value_name="gdp_millions")
).dropna()

tourism_labels = { # labels of industries considered to be tourism related
    " GDP: Arts, Entertainment, and Recreation  ($Mil)",
    " GDP: Accommodation and Food Services  ($Mil)"
}

```

```

df_long["sector_group"] = df_long["series"].apply( # function to group tourism_
    ↪related industries
    lambda s: "Tourism" if s in tourism_labels else s
)

df_grouped = (
    df_long
    .groupby(["quarter", "sector_group"], as_index=False).sum()
)

fig = px.line( # plot all sectors
    df_grouped,
    x="quarter",
    y="gdp_millions",
    color="sector_group",
    title="Hawaii GDP by Industry",
)
fig.update_layout(
    xaxis_title="Quarter",
    yaxis_title="GDP (Millions USD)",
    legend=dict(
        title="Sector Group",
    )
)
fig.show(renderer="notebook")

```

It can be seen that the tourism group makes up the third most of Hawaii's GDP.

After interpreting my findings, I came to the conclusion that Hawaii's economy does infact rely heavily on tourism. It could be seen that Hawaii has consistently made up about 8-12% of Hawaii's economy from 2005 to 2024, which is huge margin of the state's GDP. Not only that, but between all the other industries, these two industries combined make up the third most of Hawaii's total GDP. The one thing that is concerning, however, is how in 2020 during the COVID-19 pandemic, a huge drop occurred in the tourism related industries. Which leads me to my next research question.

### 1.6.2 What economic trends or declines did Hawaii experience during the COVID-19 pandemic when tourism was severely restricted?

To answer this question, I first defined the COVID-19 pandemic period as starting in 2020 Q1 and ending in 2022 Q1. I then analyzed three key metrics across this timeframe:

- **Visitor Arrivals** (as a measure of tourism)
- **Employed Civilians** (as an indicator of unemployment)
- **GDP from Accommodation and Food Services & Arts, Entertainment, and Recreation** (as a tourism-dependent industries of the economy)

These metrics were filtered to include only statewide data to reflect Hawaii's overall economic

trends. I merged together these datasets and visualized them in a multi-line plot with individual Y-axis scaling to show how they varied across different units.

```
[8]: def melt_dataframe(df, id_vars, var_name="quarter", value_name="value"):
    """
    Given a DataFrame, the columns to keep fixed, the name for the new variable,
    a name for the new value column, melts a wide-format DataFrame into
    long-format for time series data.

    If var_name not provided, defaults to 'quarter'
    If value_name not provided, defaults to 'value'

    Returns the melted long-format DataFrame.
    """
    return df.reset_index().melt(id_vars=id_vars, var_name=var_name,
    value_name=value_name)

tourism_data_long = melt_dataframe( # melt tourism data
    tourism_data,
    ["indicator", "destination"]
)
employment_data_long = melt_dataframe( # melt employment data
    employment_data,
    ["indicator", "area"]
)
gdp_data_long = melt_dataframe( # melt employment data
    gdp_data,
    ["series"]
)

tourism_filtered = tourism_data_long[ # filter tourism for arrivals statewide
    (tourism_data_long["indicator"] == "Visitor arrivals") &
    (tourism_data_long["destination"] == "Statewide")
]

employment_filtered = employment_data_long[ # filter employment for employed
    civilians statewide
    (employment_data_long["indicator"] == "Employed, Civilian") &
    (employment_data_long["area"] == "State of Hawaii")
]

gdp_filtered = gdp_data_long[
    (gdp_data_long["series"] == " GDP: Accommodation and Food Services ")
    ($Mil) |
```

```

    (gdp_data_long["series"] == " GDP: Arts, Entertainment, and Recreation _
    ↳($Mil)")
]

gdp_filtered = gdp_filtered.groupby("quarter")["value"].sum().reset_index() #_
    ↳sum gdp across tourism industries

merged_df = tourism_filtered.merge( # merge together all three filtered datasets
    employment_filtered, on="quarter", how="inner", suffixes=("_tourism",_
    ↳"_employment")
).merge(
    gdp_filtered, on="quarter"
)

merged_df.rename(columns={ # rename columns for better graph visualization
    "value_tourism": "Visitor Arrivals",
    "value_employment": "Employed Civilians",
    "value": "Tourism GDP"
}, inplace=True)

merged_df = merged_df[["quarter", "Visitor Arrivals", "Employed Civilians",_
    ↳"Tourism GDP"]] # pick out columns
merged_df = merged_df[ # filter for quarters between 2020 Q1 and 2022 Q2
    (merged_df["quarter"] >= "2020_q1") & (merged_df["quarter"] <= "2022_q1")
]

merged_df = merged_df.melt( # melt merged DataFrame for easier plotting
    id_vars=["quarter"],
    value_vars=["Visitor Arrivals", "Employed Civilians", "Tourism GDP"],
    var_name="Metric",
    value_name="Value"
)

fig = px.line( # plot multiline plot
    merged_df,
    x="quarter",
    y="Value",
    color="Metric",
    title="Metrics from 2020 Q1 to 2022 Q1",
    markers=True,
    facet_row="Metric",
    log_y=True
)

fig.update_layout(
    xaxis_title="Quarter",

```

```

        yaxis_title="Value",
    )
    fig.update_yaxes(type="log") # change to logarithmic scale

    fig.for_each_annotation(lambda a: a.update(text=a.text.split("=")[-1])) #
        ↪ create y-axis notations

    fig.for_each_yaxis(lambda yaxis: yaxis.update(matches=None)) # dont standardize
        ↪ fixed y-axes across plots

    fig.show(renderer="notebook")

# test melt_dataframe #
df = pd.DataFrame({
    "indicator": ["A", "B"],
    "destination": ["X", "Y"],
    "2020_q1": [10, 20],
    "2020_q2": [15, 25]
})

melted = melt_dataframe(df, id_vars=["indicator", "destination"])

all_cols = list(df.reset_index().columns)
value_cols = [c for c in all_cols if c not in ["indicator", "destination"]]
expected_rows = df.shape[0] * len(value_cols)

assert len(melted) == expected_rows, "Incorrect row count"
assert set(melted["quarter"].unique()) == set(value_cols), "Quarter doesn't
    ↪ have the right columns"

val = melted.loc[
    (melted["indicator"] == "B") & (melted["quarter"] == "2020_q2"),
    "value"
].values[0]
assert val == 25, "Values don't match up"

```

Upon analyzing the graph, it could be seen that Visitor Arrivals saw a sharp decline in 2020 Q2, dropping dramatically from over 2 million to under 100,000 due to lockdowns and travel restrictions. Unemployment also surged during this time, with employment dropping to its lowest of 525k in 2020 Q2, reflecting the collapse in tourism-related jobs. Tourism-related GDP also declined during this period, though more gradually than arrivals, showing economic hardship in tourism-reliant sectors. As tourism began recovering in late 2020 and into 2021, improvements in both GDP and employment followed, though not immediately returning to pre-pandemic levels by 2022 Q1.

This demonstrates the direct impact of tourism restrictions on both employment and economic output. It shows how as a result of visitor arrivals dropping during the pandemic in 2020 Q2, the number of civilians had dropped, as well as the GDP from tourism-related industries.

### 1.6.3 How did the decline in visitor arrivals during the COVID-19 pandemic affect state-level housing affordability?

Residents often claim that tourists coming to Hawaii oftentimes inflates housing prices, most likely due to affording long/short-term rentals in Hawaii during their stay and even taking up the already relatively low amounts of housing available to residents. Due to this demand for housing in Hawaii, prices get raised higher and higher every year. For this reason, there's a substantial majority of homeless residents in Hawaii, with 80.5 out of 10,000 people being homeless (<https://usafacts.org/articles/which-states-have-the-highest-and-lowest-rates-of-homelessness/>).

To answer this question on whether or not tourists inflate housing prices, I planned on using another dataset regarding median housing prices across certain counties of Hawaii, and visitor arrivals in Hawaii, then comparing these two datasets on a dual-axis time series line chart. The only downside to the dataset of median housing prices is that it only contained quarterly data for Honolulu and yearly data for Maui. Visitor arrivals are also only counted statewide, which in turn will mask the housing rates across the different islands. To address this issue, I decided to focus solely on housing prices in Honolulu inside my housing prices dataset. I justify this usage of comparing Honolulu's housing prices to housing prices as Oahu (considered Honolulu county) is one of the most visited islands throughout Hawaii, accounting for 52% of visitors across the state in 2024 alone (<https://www.hawaii-guide.com/hawaii-tourism-statistics>).

```
[9]: # melt the two DataFrames using my previous melt_dataframe function
housing_long = melt_dataframe( # melt housing
    housing_data,
    id_vars=["indicator", "area"],
    value_name="median_price"
)

tourism_long = melt_dataframe( # melt tourism
    tourism_data,
    id_vars=["indicator", "destination"],
    value_name="visitor_arrivals"
)

housing_filtered = housing_long[ # filter housing in honolulu
    (housing_long["indicator"] == "Single Family Home- Median Price") &
    (housing_long["area"] == "Honolulu County")
]

tourism_filtered = tourism_long[ # filter statewide visitor arrivals
    (tourism_long["indicator"] == "Visitor arrivals") &
    (tourism_long["destination"] == "Statewide")
]

merged_df = pd.merge( # merge together the two datasets
    tourism_filtered[["quarter", "visitor_arrivals"]],
    housing_filtered[["quarter", "median_price"]],
    on="quarter",
```

```

        how="inner"
    ).sort_values("quarter")

fig = go.Figure() # plot

fig.add_trace(go.Scatter( # plot first line of visitor arrivals
    x=merged_df["quarter"],
    y=merged_df["visitor_arrivals"],
    name="Visitor Arrivals",
    yaxis="y1",
    mode="lines+markers"
))

fig.add_trace(go.Scatter( # plot second line of housing prices
    x=merged_df["quarter"],
    y=merged_df["median_price"],
    name="Median Housing Price (Honolulu)",
    yaxis="y2",
    mode="lines+markers"
))

fig.update_layout(
    title="Visitor Arrivals vs. Honolulu Median Housing Price (COVID-19_
↪Impact)",
    xaxis_title="Quarter",
    yaxis=dict(
        title="Visitor Arrivals",
        side="left"
    ),
    yaxis2=dict(
        title="Median Price (USD)",
        overlaying="y",
        side="right"
    ),
)

fig.show(renderer="notebook")

```

Analyzing the plot, it's observable that there exists a slight correlation between visitor arrivals and median housing price. Starting from 2005 Q2, it can be seen that the two variables follow the same trend, one increasing as the other increases. This correlation is lost when the pandemic hits in 2020 Q2 and tourism is restricted, where the median housing price rapidly increases. Then after tourism is opened up and visitors come to Hawaii, the Median Housing Price keeps correlation with visitor arrivals as well.

From this, it can be seen that visitor arrivals may have some sort of association/correlation with the increase in the cost of housing. As the number of visitors rise, so does the median housing price. This may likely be from broader economic growth. However, this correlation does not necessarily

prove causation, as the median of the housing price continues to increase as the number of visitors increases. Thus, tourism is not a direct cause of housing price inflation, at least from this graph. This housing price inflation could be caused by other demographic or economic factors.

## **1.7 Implications and Limitations**

This analysis highlights Hawaii’s economic dependence on tourism and explores how this dependence affects the state’s overall well-being, such as employment and housing. Policymakers could benefit from these findings to establish funding and advocacy in alternative industries for a more diverse economy outside of just tourism. They could also look at insights on how this drop in tourism caused large surges in unemployment and create safety nets with stimulus checks during these suffrages. economic development organizations could benefit from this analysis by investing into other industries which could diversify and give independence to Hawaii’s economy. The general public could benefit from this analysis by reassessing their claims on how tourism affects housing. By revealing tourism’s influence on GDP and employment and its possible correlation with housing affordability, this project supports informed decision-making toward a more independent economy for Hawaii, as well as more general knowledge on the effects of tourism on Hawaii’s general well-being as well.

Some limitations include:

### **1. Causation vs. Correlation**

While the data shows a correlation between tourism activity and median housing prices, this does not mean causation. Housing costs are shaped by many factors beyond visitor arrivals that are within the breadth of the other economic aspects. The graph shows these trends rising together, but that alone is insufficient to conclude that tourism causes housing inflation.

### **2. Incomplete Geographic and Temporal Coverage**

The analysis on housing affordability uses data only from Honolulu County, while visitor arrival data is statewide. Oahu is the most visited island, which somewhat justifies this comparison, but it doesn’t account for variation in housing markets in other islands (i.e. Maui, Kauai, Big Island).

### **3. Pandemic-Specific Disruptions**

Much of the analysis centers on trends during the COVID-19 pandemic. Economic behavior during this time might not reflect standard market responses. For instance, the drop in employment and GDP was probably influenced by other factors like government mandates, international policy, and supply chain issues that go are beyond Hawaii’s tourism industry.

Because of these limitations,, the findings should not be used to argue for or against tourism without deeper exploration. Rather, this analysis serves as a starting point for future analyses, encouraging Hawaii to examine both the benefits and vulnerabilities of tourism in a broader economic viewpoint.