

Forecasting Japan's Cherry Blossom Dates

Done by: Amanda Luah
Date: 5 December 2024

TABLE OF CONTENT

Problem, Goals, and Audiences.....	3 - 9
Problem, Goals, and Audiences.....	3 - 6
Criteria for success and Overview of Approach.....	7 - 9
Data Sources and Definitions in a Data Dictionary.....	10
Patterns, Trends and Insights	11 - 22
Data source and Data cleaning/processing.....	11 - 12
Exploratory Data Analysis (EDA).....	13 - 21
Interactive Dashboard.....	22
Predictive Model.....	23 - 25
Recommendations.....	26 - 29
Mr Yashida Noboru.....	26
Mdm Misaki Chiyu.....	27-28
Mr Takai Tsukasa.....	29

Problem, Goals and Audiences

Problem statement


How can we help local businesses to maximize their success during the cherry blossom period?

Goal


Develop a reliable prediction model for Cherry Blossom dates across various cities in Japan in 2024 & 2025.


Target Audience

We have 3 target audiences mainly Business Owners whose businesses are affected by the Cherry Blossom and require our prediction model to solve their seasonal challenge. Please refer to the tables for their full profile:

Business owner #1	
Portrait	
Name:	Yashida Noboru
Business Type:	Ryokan (Traditional Inn) Owner
Location:	Kyoto

Goals:	<ul style="list-style-type: none"> • Wants to attract more tourists during Cherry Blossom season with special packages
Challenges:	<ul style="list-style-type: none"> • Variability in bloom dates leads to difficulty in planning peak season promotions • Not enough bookings during cherry blossom season
Needs:	<ul style="list-style-type: none"> • Reliable cherry blossom forecasts for marketing and logistics • Early forecasts to prepare special packages and staff
How our forecast may help:	<ul style="list-style-type: none"> • Provides precise bloom predictions to schedule peak-season offers and tours. • Allows her to plan promotional campaigns around cherry blossom dates

Business owner #2	
Portrait	
Name:	Misaki Chiyu
Business Type:	Local Souvenir Shop Owner - Specialty in local alcohol and snacks
Location:	Kochi
Goals:	<ul style="list-style-type: none"> • Wants to maximize foot traffic during Cherry Blossom season • Wants to ensure new inventory is enough for the upcoming season
Challenges:	<ul style="list-style-type: none"> • Inventory obsolescence due to poor business • Missed sales opportunities due to inaccurate promotion timing
Needs:	<ul style="list-style-type: none"> • Reliable forecasts so that she can estimate when tourists will flock to the region
How our forecast may help:	<ul style="list-style-type: none"> • Helps with targeted promotional timing (e.g., discounts on sakura-themed items).

Business owner #3	
Portrait	
Name:	Takai Tsukasa
Business Type:	Local Restaurant Chef Owner - Izakaya
Location:	Fukuoka
Goals:	<ul style="list-style-type: none"> • Create seasonal menus to attract customers • Capitalize on the increase in foot traffic during cherry blossom season
Challenges:	<ul style="list-style-type: none"> • Timing seasonal menus can be tricky if bloom dates are uncertain • Short bloom periods limit the marketing window
Needs:	<ul style="list-style-type: none"> • Early forecasts to plan ingredient procurement and marketing
How our forecast may help:	<ul style="list-style-type: none"> • Helps launch seasonal items at the right time to maximize sales • Improves customer satisfaction by syncing offerings with the bloom period.

Criteria for success

- ❖ Forecasted date \pm 3 days of actual blooming date
- ❖ Being able to predict blooming dates across most region of Japan
- ❖ Test the model's accuracy on unseen data

Overview approach of solving our problem

Forecasting cherry blossom dates is a critical task for local businesses in Japan to plan events, promotions, and logistics during the sakura season.

Here's a structured overview of the approach to solving this problem:

1. Context

Cherry Blossoms in Japan usually bloom in a linear fashion, moving from the South to the North between mid-March and early May, with different regions blooming at different times. Cherry blossoms typically last for 10 to 14 days, with only certain species lasting longer than 14 days. Hence, it is very important to accurately predict the first bloom and for businesses to make use of the 14 days to improve their business.

2. Understanding the problem

We saw that there is an ongoing issue of accuracy for Cherry Blossom forecasting in Japan, which affects many people including Local Businesses, Tourism Boards, Event Planners and Travellers. With this in mind, we will solely focus on developing a reliable prediction model for Local Businesses in various cities of Japan and provide actionable insights for our stakeholders.

3. Data Preparation

To create our prediction mode, we will need to source for historical data of Cherry Blossom bloom dates (preferably in the past 10 years or more) as well as Weather data like Monthly Average Air Temperature for March & April. As the Weather data is not readily available from any open source like Kaggle, we will retrieve the data from Japan Meteorological Agency (an official Government Agency dedicated to scientific observation and research of natural phenomena).

For our testing data, we will retrieve the 2024 data of Air Temperatures as well as location information like latitude, longitude etc for various regions of Japan from Kaggle.

Once all data has been retrieved, we will proceed with data cleaning such as handling missing or inconsistent entries, standardizing all temperature and date format, as well as aligning all regional bloom data with respective weather data. Along the way, we will create additional columns if necessary.

4. Modeling Approach

a. Exploratory Data Analysis (EDA)

Using our sourced data set, we will visualize the trends of blooming dates over the years across regions, investigate correlations between different variables as well as any other interesting patterns/trends that can be explored.

b. Model Selection

As our intention focuses on predicting future blooming dates, we will be using the Time Series model - ARIMAX (Auto-Regressive Integrated Moving Average with Exogenous Variables) to incorporate time series and temperature trends.

c. Regional modelling

As our results vary by region and climatic changes are different for each region, our model will be trained separately for every region and forecasted by region as well.

5. Forecasting

The trained models will be used to predict blooming dates for 2024 based on the weather data as well as predict future blooming dates without any current data.

6. Strategic Recommendations for Businesses

With our model, we can find patterns/trends to determine peak bloom weeks for event promotions, recommend best periods for marketing campaigns as well as monitor real-time weather changes and update forecasts to business operations dynamically.

7. Continuous Improvement

The model can be updated periodically with new data to improve accuracy and narrow down the range.

8. Tools and Technologies

Data Cleaning, EDA, Modelling: Python, Jupyter

Data Visualisation: Tableau, Python, Jupyter

Data Sources and Definitions in a Data Dictionary

Train Data

Field	Type	Description	Reference
Index	Integer	Reference number	
Region	Text	Regions of Japan	
Year	Integer	Year of bloom date	
Bloom_date	Date	Historical bloom dates	
Mar_avg_temp	Numeric	Average air temperature in March	
Apr_avg_temp	Numeric	Average air temperature in April	

Test Data

Field	Type	Description	Reference
Region	Text	Regions of Japan	
Year	Integer	Year of bloom date	
Mar_avg_temp	Numeric	Average air temperature in March	
Apr_avg_temp	Numeric	Average air temperature in April	

Patterns, Trends and Insights

1. Data source and Data cleaning/processing

There were a few datasets sourced for this project. As most of these datasets were reformatted and abstractedly taken from other open source as well, there was not much data cleaning to do for most. The following are our steps taken:

Historical blooming dates from 1990-2023

Firstly, we retrieved the historical blooming dates dataset from Kaggle but realised it was not complete in the way we wanted it to be. The columns included were Historical years and Regions. Historical years ranged from 1953 to 2023 but we decided to cut the range down to only 1990 to 2023 as data spanning from too long ago might not be very useful when taking into consideration the weather that is ever changing. There were also over 100 regions in the original dataset but only 25 were selected to be included in the final dataset.

In the Jupyter notebook, we checked for any null values and dropped duplicates, if any. We then converted the date column format to standard calendar format YYYY-MM-DD so that we could easily convert these dates to Julian days. Julian days was an additional column added to the dataset, since our prediction model needed Julian days as a y-variable. We then exported the cleaned data to csv and analysed the data for EDA purposes.

Source:

https://www.kaggle.com/datasets/ryanglasnapp/japanese-cherry-blossom-data/dataset?select=sakura_first_bloom_dates.csv

https://www.data.jma.go.jp/obd/stats/etrn/view/monthly_s3_en.php?block_no=47588&view=1

Cherry blossom locations

The second dataset was also sourced from Kaggle and included latitude and longitude information about various regions of Japan with specific cherry blossom spots included. This dataset was extremely useful as it helped in the creation of geographical maps as well as other charts/graphs for visualization purposes.

When trying to combine the two dataset together, we realised that there was a limitation because the number of prefectures for both datasets do not match each

other, hence we could only include common prefectures between the two as the regions we would forecast the blooming dates for. Therefore, only 25 regions were selected as mentioned above.

Source:

kaggle.com/datasets/altabbt/japan-cherry-blossoms-forecasts-2024?select=cherry_blossom_forecasts.csv

2024 Weather data

The last dataset was mainly retrieved from the Japan's Meteorological Agency's webpage and was meant to be used as training data for our prediction model. The columns included were Region, Year, Mar_avg_temp and Apr_avg_temp. We took the 25 regions selected from the Historical blooming dates dataset and made it consistent for this dataset. The respective average air temperatures for March and April 2024 were taken from Japan's Meteorological Agency's webpage.

Source:

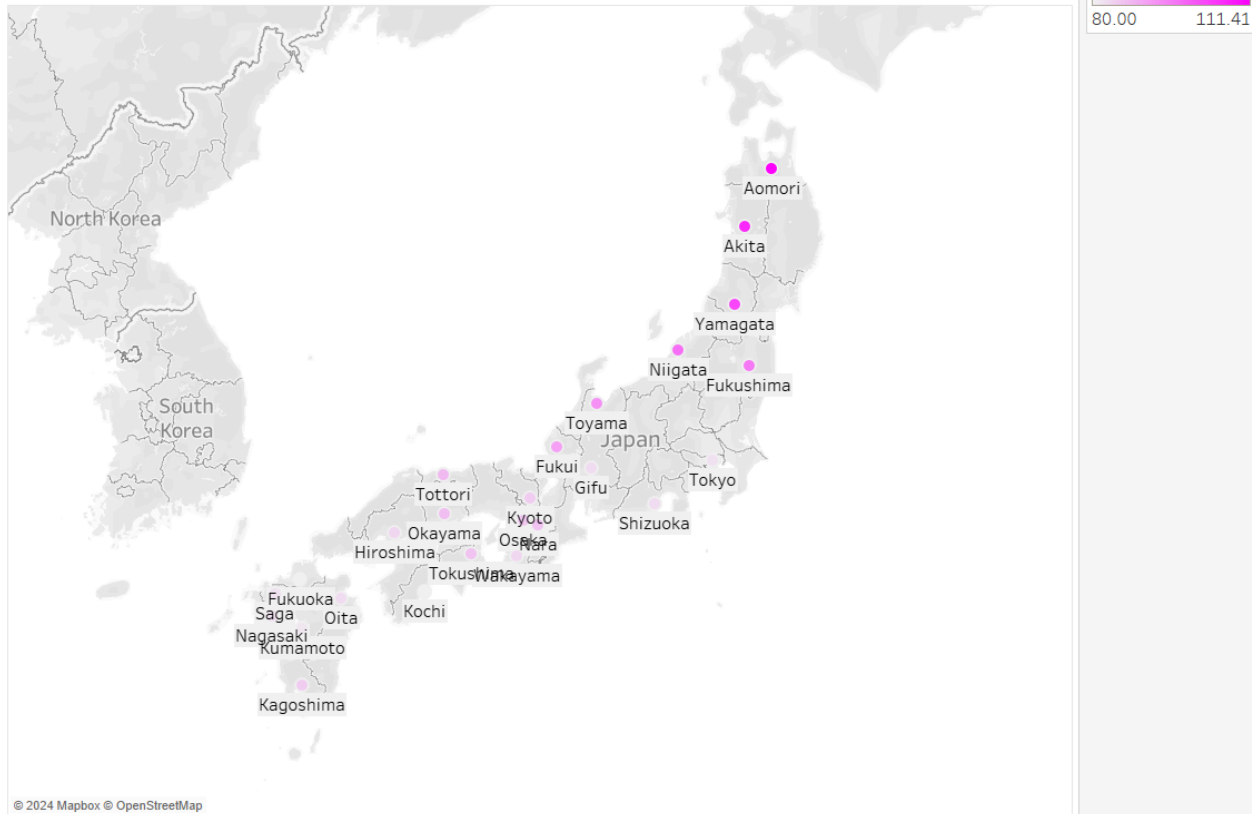
https://www.data.jma.go.jp/obd/stats/etrn/view/monthly_s3_en.php?block_no=47588&view=1

2. Exploratory Data Analysis (EDA)

We created the following visualizations in [Tableau](#) for our exploration of the data:

- Blooming dates across different parts of Japan (Geographical Map)

Blooming dates across different parts of Japan

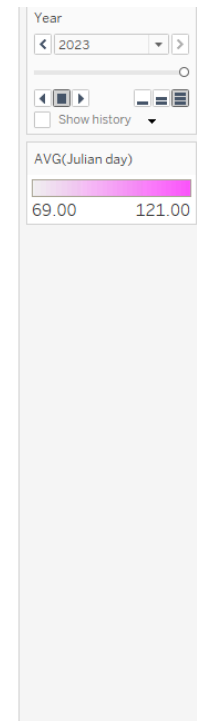
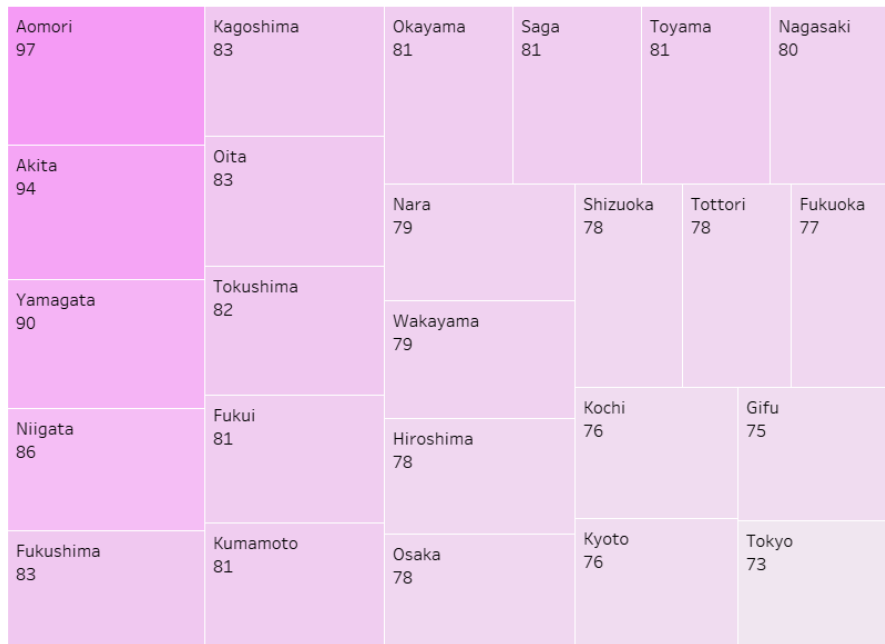


Context: We have created a Geographical Map for easy visualization across the map on the average bloom dates by Year, with the darkest shade of pink being late blooming cities and the lightest shade of pink being early blooming cities.

Findings: Aomori is consistently the city with the highest average bloom date (Julian Day) from 1990-2023, which means amongst all the cities included in this model, Aomori is the last city to have their first bloom.

- Average bloom dates by year (Heat Map)

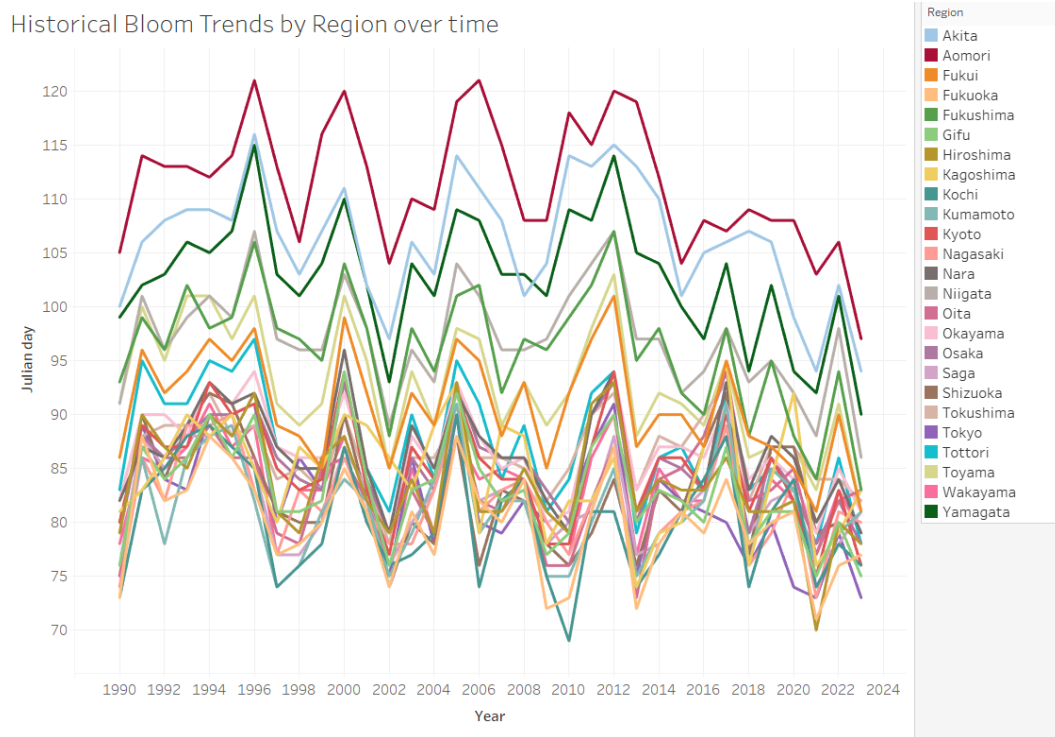
Average Bloom Dates by Year - 2023



Context: We have created a Heat Map to visualize the average bloom dates by Year, with the darkest shade of pink being late blooming cities and the lightest shade of pink being early blooming cities.

Findings: In line with the geographical map, Aomori is consistently the city with the highest average bloom dates from 1990 to 2023 which means it has always been a late blooming city due to it being in the Tokuho region and is one of the last few regions to bloom.

- Historical bloom trends by cities over time (Line Graph)

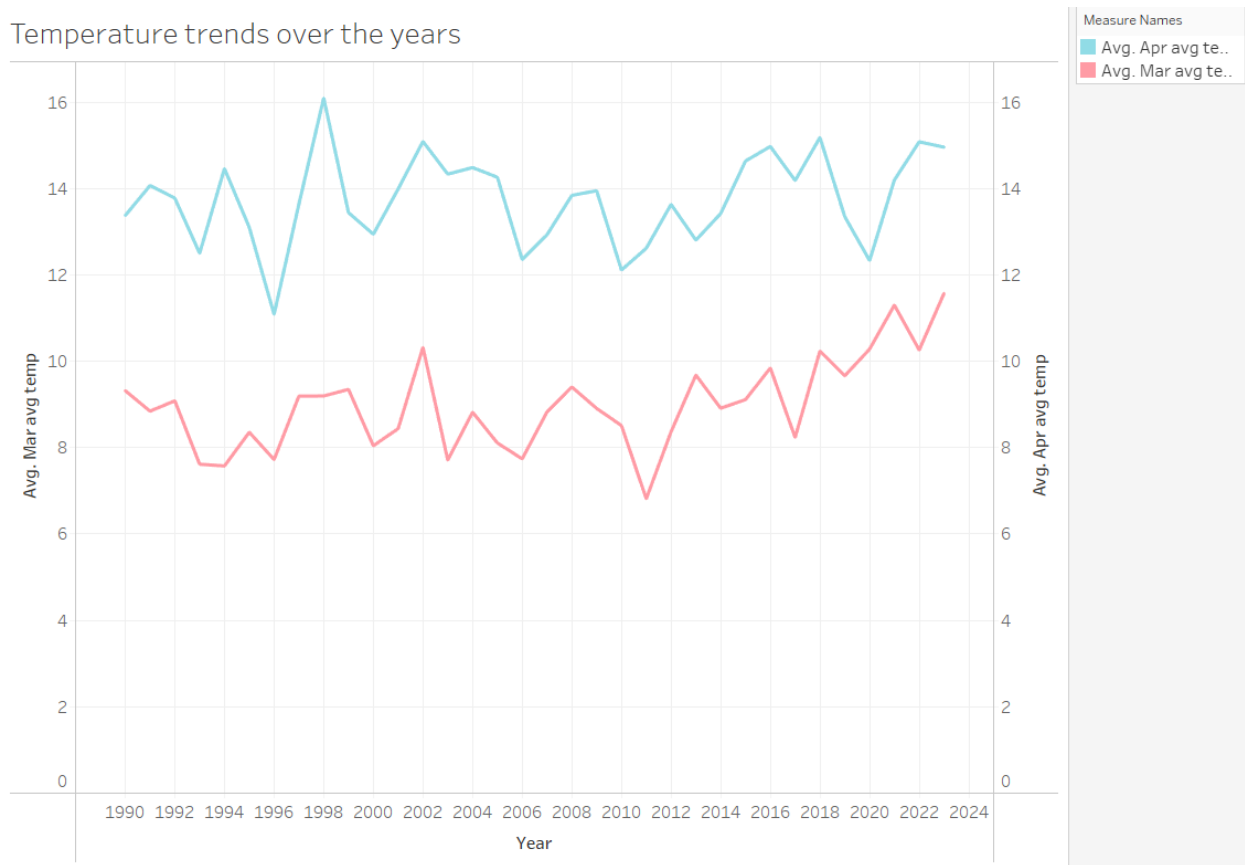


Context: We have created a Line graph to visualize the historical bloom trends by cities over time, with the different colours representing the 25 regions.

Findings: In line with our previous findings, Aomori is consistently the city with the highest average Julian day from 1990 to 2023 which means it has always been a late blooming city due to it being in the Tokuho region and is one of the last few regions to bloom.

- Temperature trend over the years (Line graph)

Temperature trends over the years

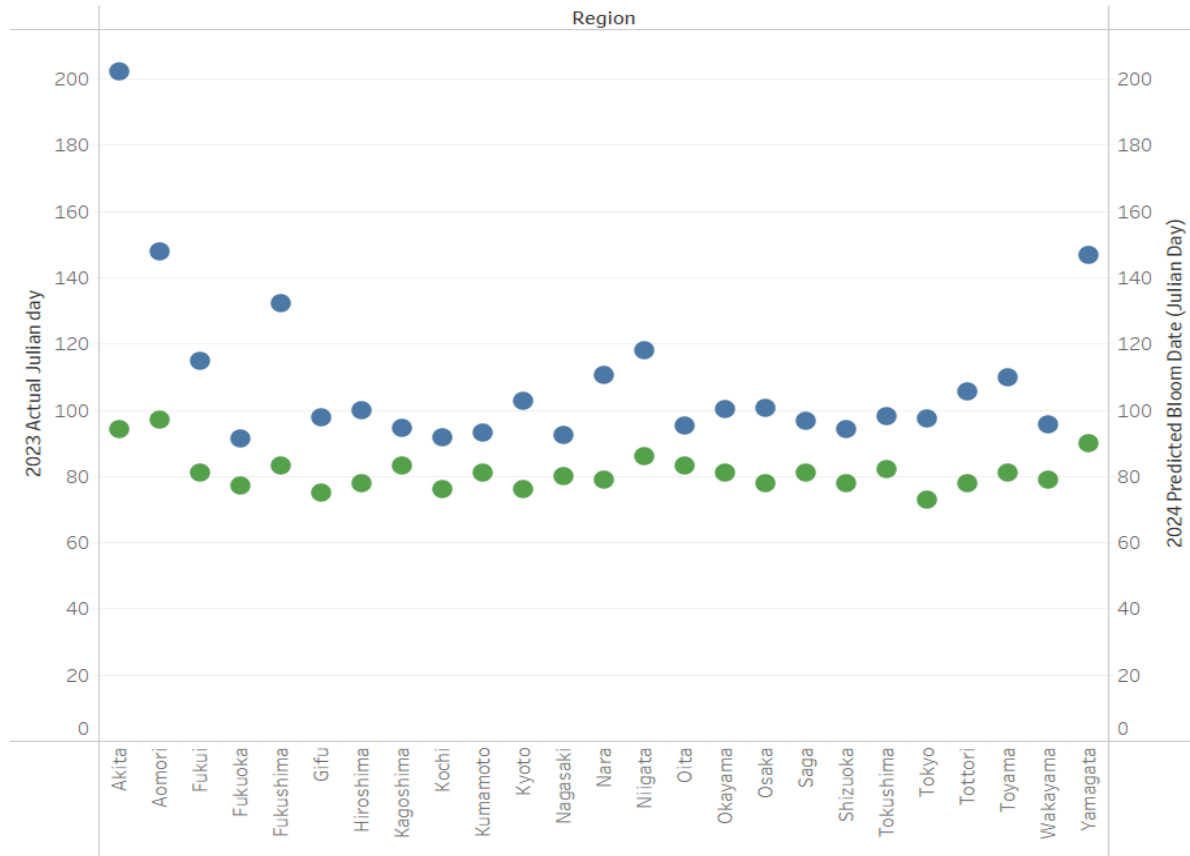


Context: We have created a Line graph to visualize the March and April's average air temperature over the years, with pink representing March and Blue representing April.

Findings: We can confirm that the weather is getting warmer, with both March and April's temperature increasing over the years, which in general causes cherry blossoms to bloom earlier.

● 2023 Actual vs 2024 Predicted Blooms

2023 vs 2024 Predicted Bloom in Julian Days

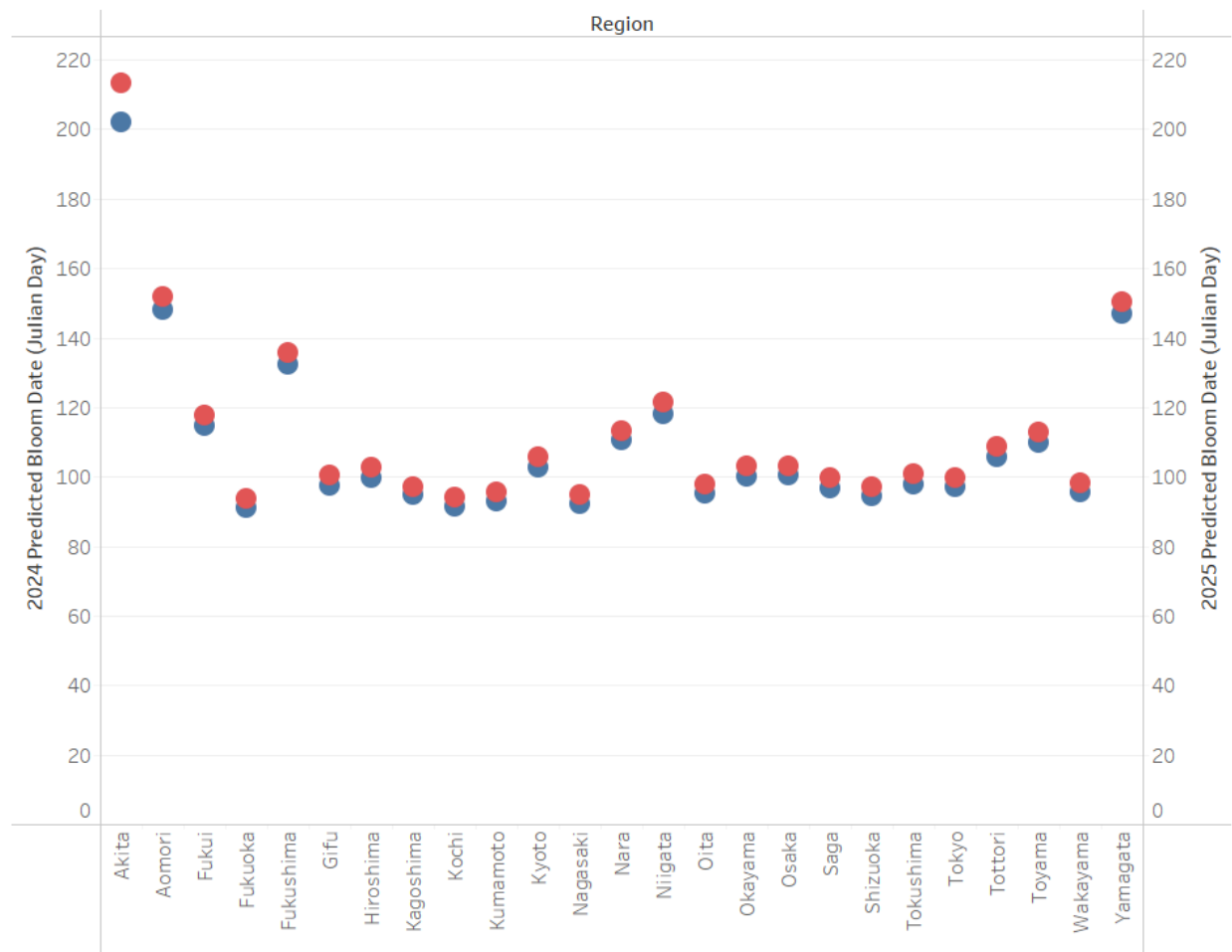


Context: We have created a scatter plot graph to visualize the blooms in Julian days between 2023 actual and 2024 predictions by our model, with green representing 2023 and blue representing 2024.

Findings: The bloom for 2024 is predicted to be much later than those in 2023, which is true after researching online. For Tokyo, cherry blossoms bloomed 15 days later compared to 2023.

- 2024 vs 2025 Predicted Blooms

2024 vs 2025 Predicted Bloom in Julian Days

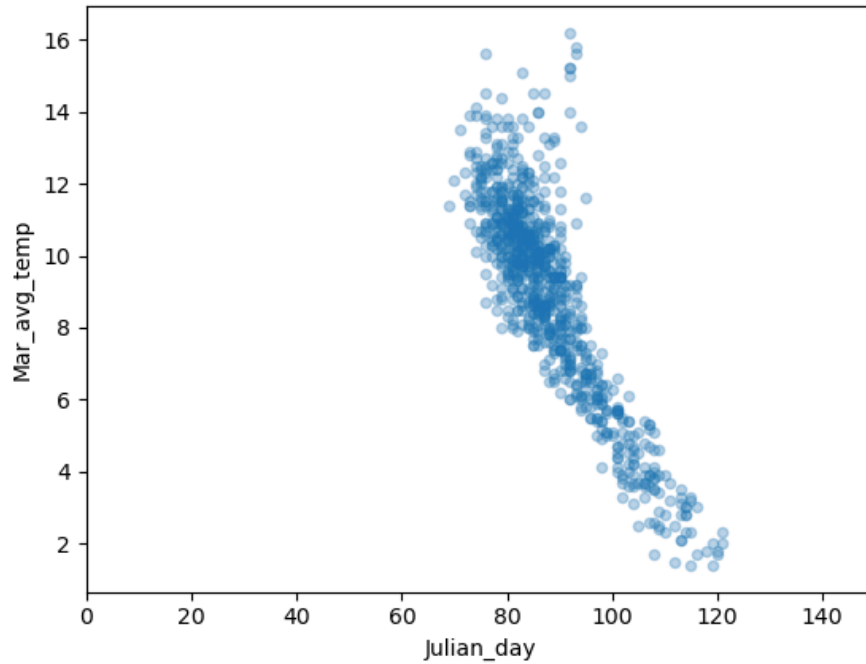


Context: We have created a scatter plot graph to visualize the predicted blooms in Julian days between 2024 and 2025 predictions by our model, with blue representing 2024 and red representing 2025.

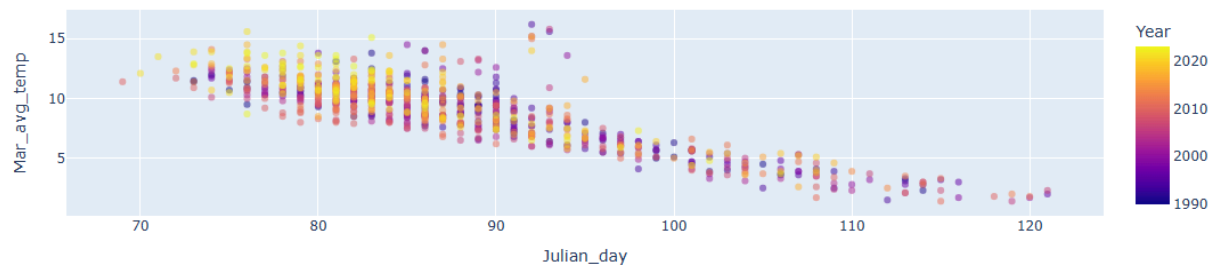
Findings: The pattern is quite similar because 2025's predictions were outputted using the data from 2024, with the exception of Julian days getting higher, which means blooming gets later. This seems to be similar to our previous graph when comparing 2023 and 2024.

We created the following visualizations in our Jupyter Notebook for our exploration of the data:

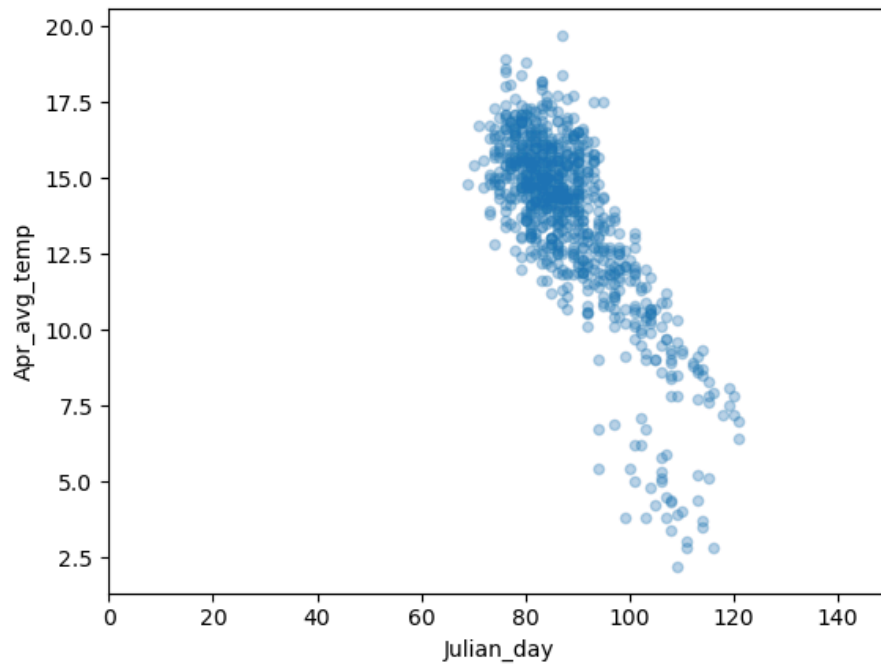
- Scatter plot for relationship between March Average Temperature and Bloom date



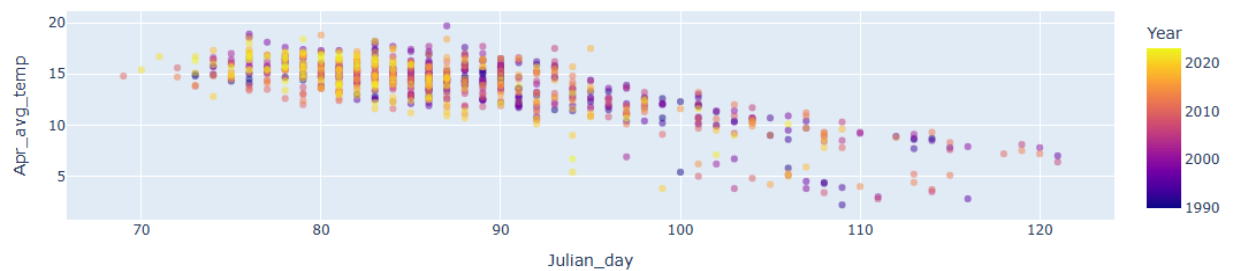
March Temperature vs Bloom Date



- Scatter plot for relationship between April Average Temperature and Bloom date

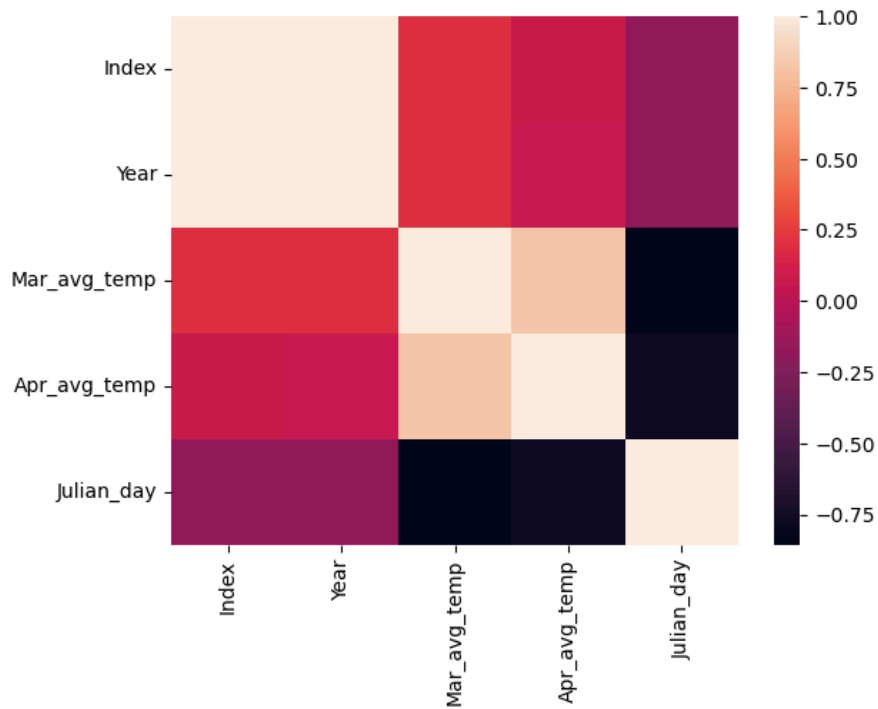


April Temperature vs Bloom Date



Finding: From the scatterplots, we could tell that the higher the air temperature, the lower the Julian day, which means the earlier the bloom. We can also tell that the air temperature has been getting warmer by year, which means the first bloom tends to be earlier as the year goes by.

- Correlation matrix heatmap

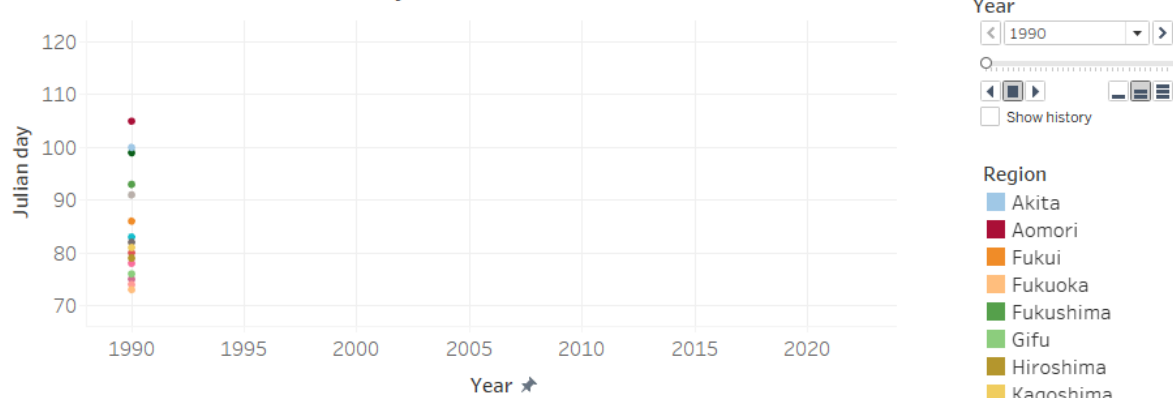


Finding: From the correlation heatmap, we could tell that March and April's Average air temperature & Julian day were the most strongly correlated.

3. Interactive Dashboard components

We created an interactive dashboard on Tableau - consisting of historical bloom trends by cities over time and average Julian days by year, which helps our targeted audience to look at historical bloom trends over time, with the year and city easily filterable on the visuals for easy reference. This is so that even if they do not know the exact bloom date for the next year, they are able to look at the pattern at an overview.

Historical Bloom Trends by cities over time - 1990



Average Bloom Dates (Julian days) by Year - 1990

Aomori 105	Toyama 91		Kyoto 80	Osaka 80	
Akita 100	Fukui 86	Hiroshima 79	Gifu 76		Saga 76
Yamagata 99	Tottori 83	Shizuoka 79			
Fukushima 93	Nara 82	Tokyo 79	Oita 75		
Niigata 91	Okayama 82	Wakayama 78	Nagasaki 74		

Predictive Model

The ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) model is a time-series forecasting technique that predicts outcomes based on:

- **Historical data patterns** (trend and seasonality)
- **External factors (exogenous variables)** that influence the outcome

In our case:

- **Historical Data:** Past cherry blossom bloom dates (Julian Day format).
- **External Variables:** Average temperatures for March and April.

Why do we use ARIMAX for Cherry Blossom Forecasting?

Cherry blossom blooming is influenced by:

- **Temperature patterns:** Warmer springs can cause earlier blooming.
- **Yearly trends:** Historical data reveals recurring seasonal patterns and long-term shifts. ARIMAX allows us to incorporate both.
- **Time-series trends** (using past bloom dates).
- **Environmental factors** (using March and April Air Temperatures)

How Does ARIMAX Work?

1. **AutoRegressive (AR) Component:**
 - Captures how past bloom dates affect current bloom dates.
 - For example, if blooms were earlier in recent years, this pattern may continue.
2. **Integrated (I) Component:**
 - Accounts for long-term trends, such as blooming dates shifting earlier due to climate change.
3. **Moving Average (MA) Component:**
 - Models short-term fluctuations in blooming dates caused by unusual weather events.

4. **Exogenous Variables (X):**

- Adds temperature data (March and April averages) to improve accuracy.
- Warmer March/April temperatures are linked to earlier blooming.

Inputs to the Model

- **Julian Day:** The day of the year when cherry blossoms bloom (e.g. March 31 = Julian Day 90).
- **Year:** Historical timeline for analyzing long-term trends.
- **March & April Average Temperatures:** Key environmental factors influencing bloom dates.

What Insights Does the Model Provide?

- **Bloom Date Forecast:** Predicts the most likely bloom date for each city in 2024 and 2025.
- **Confidence Intervals:** Provides a range of dates within which blooming is highly likely.

Model Validation and Accuracy

- The model is trained using **historical data**.
- The model is tested using **current 2024 data**.
- Validation ensures the model accurately captures:
 - Trends (e.g., earlier blooming over decades).
 - Seasonal effects (e.g., typical March-April bloom window).
- Performance is evaluated using metrics like **Mean Absolute Error (MAE)**

Output

Mean Absolute Error (MAE) = 16.019

Outcome

The model was not able to predict the 2024 bloom dates as accurately as possible, most likely because we looked at Air Temperature as the only variable involved. The training dataset itself also had limitations such as using the monthly average air temperature instead of daily average air temperature when training the model. The Japan Meteorological Agency only provided the monthly mean air temperature for each region, hence that was the only data we had.

Another limitation we had was the number of months of weather data that we had, only March and April. It is essential to know that months preceding the Spring season were important as well. These flowers needed cold weather in fall and winter to wake up in the spring. When spring comes and it gets warmer, the buds bloom. In places with warm winters, cherry trees actually bloom later.

Recommendations

For Mr Yashida Noboru (Ryokan (Traditional Inn) Owner) in Kyoto, our recommendations would be:

1. Address bloom variability

- Use our cherry blossom forecasting model to predict peak bloom dates for the Ryokan's location.
- Plan promotions with flexibility, offering “cherry blossom week” packages that can be adjusted based on real-time bloom updates.
- Highlight the full cherry blossom experience, not just peak bloom.
- Offer photo tours or guided walks to nearby cherry blossom viewing spots throughout the season.

2. Increase bookings

- Encourage early bookings by offering early bird discounts for reservations made 2-3 months before the anticipated season.
- Provide flexible cancellation and refund policies tied to bloom variability, ensuring potential guests to feel secure booking early.
- Provide additional perks for weekday stays such as free breakfast & weekday stay discounts.

3. Marketing strategies

- Partner with regional cherry blossom festivals to cross-promote his services
- Leverage social media to post real-time updates about the blossoms near his inn.
- Focus on domestic travellers and tourists from nearby cities
- Utilize online travel platforms (e.g. Agoda) to highlight proximity to famous cherry blossom spots.
- Host cherry blossom viewing picnics with seasonal bentos
- Provide guests with unique keepsakes like sakura-scented candles/soaps or crafts.
- Take in feedback from guests to understand what they enjoy most and where improvements are needed.

For Mdm Misaki Chiyu (Local Souvenir Shop Owner - Specialty in local alcohol and snacks) in Kochi, our recommendations would be:

1. Maximize foot traffic during cherry blossom season

- Offer seasonal promotions on alcohol and snacks that are most suitable for cherry blossom season
- Host in-store tasting sessions featuring in season sake, beers or snacks, encouraging customers to explore seasonal offerings before committing to a purchase.
- Partner with local cherry blossom festival organizers to promote her shop as a “must-visit” destination for tourists.
- Run geo-targeted ads on platforms such as Instagram and Google to reach tourists visiting the area.
- Create special promotions such as receiving further 15% off items if customers share on social media and tag the business.

2. Ensure good inventory levels

- Use historical sales data to identify best-selling items during past cherry blossom seasons.
- Leverage cherry blossom bloom forecasts to time inventory replenishment just before peak foot traffic.
- Prioritize stocking popular seasonal products and diversify the product line to include versatile items that appeal to both tourists and locals.
- Establish partnerships with local producers to ensure exclusivity and timely delivery.

3. Reduce inventory obsolescence

- Stock items with cross-seasonal appeal, such as alcohol with general floral or fruity flavors that are not strictly tied to cherry blossoms.
- Rebrand leftover inventory as “spring specials” to extend their relevance.
- Run post-season flash deals to clear excess stock.
- Bundle slow-moving items with bestsellers to increase attractiveness.
- Offer “free gift with purchase” deals on near-expiry products.

4. Time promotions accurately

- Use our cherry blossom forecasting model to align promotions and inventory availability with expected peak tourist traffic.

- Start pre-bloom promotions 2 weeks before the predicted bloom dates to create awareness and secure early sales.

5. Enhance product appeal

- Collaborate with local breweries and confectioners to produce exclusive seasonal alcohol and snacks only available at this shop.
- Offer pre-packaged gift sets of local alcohol and snacks, mainly tailored for tourists looking for ready-made souvenirs.
- Offer customization on snack and alcohol bundles for gifting or personal use with limited edition packaging (sakura themed).

For Mr Takai Tsukasa (Local restaurant chef owner - Izakaya) in Fukuoka, our recommendations would be:

1. Seasonal Menu Planning with flexibility

- Designate a “Spring Special Menu” with exclusive dishes available only during the cherry blossom season.
- Incorporate cherry blossom or spring elements into the menu, such as sakura-infused drinks/alcohol, desserts or seasonal vegetables.
- Highlight local ingredients that align with the season, such as fresh seafood, pickled vegetables.
- Introduce a “Sakura/Spring Preview Menu” 2-3 weeks before the expected bloom to build anticipation, even if cherry blossoms have yet to bloom.
- Keep the seasonal menu flexible, allowing small adjustments based on real-time bloom dates and ingredient availability.

2. Ingredient Procurement and Inventory Management

- Use our cherry blossom forecasting model to plan ingredient procurement.
- Aside from our model, utilize services that offer long-term predictions for sakura trees, especially for predictions on how long the cherry blossom season will last.
- Try to implement reservations only in the restaurant to make it easy to anticipate the volume of ingredients needed.
- Partner with local farmers and suppliers to secure fresh, seasonal ingredients with short lead times or on-demand, to reduce waste from overstocking.
- Pre-arrange bulk purchases of seasonal ingredients ahead of time to ensure availability during the peak season.

3. Capitalizing on the short marketing window

- Start early marketing campaigns to build anticipation for the seasonal menu well before the bloom.
- Leverage on social media platforms and their website to promote the upcoming “Spring Special Menu” weeks in advance. Include teasers with visuals of planned dishes.
- Collaborate with local food bloggers to create buzz about the seasonal menu.
- Introduce reservations-only to capture premium bookings around the whole season.
- Try to market the menu as celebrating the spirit of Spring and not only the Cherry Blossoms, so as to extend the appeal beyond the short bloom period.