**I will produce project that use this dataset for topic "Data-centric product with an AI add-on
" please summary report to one poster template that include at least follwing information i.e. Title, Issue and motivation, Traditional solution, AI add-on solution, Visualization.**

Pro Search

- Analyze the provided dataset to extract relevant information for the poster template including Title, Issue and motivation, Traditional solution, AI add-on solution, and Visualization
- Summarize the findings and create a poster template with the required sections

Sources

thailand_domestic_tourism.csv
thailand_domestic_tourism

Answer

Ask follow-up
Pro

Poster: Data-Centric Product with AI Add-On for Domestic Tourism in Thailand

Title

Data-Centric Product with AI Add-On for Domestic Tourism in Thailand

Issue and Motivation

The dataset reveals the distribution of domestic and foreign tourists across various provinces in Thailand, along with their revenue contributions. The challenge lies in optimizing tourism strategies to maximize revenue while ensuring tourist satisfaction. Traditional methods of analyzing tourism data are time-consuming and prone to human error, limiting the ability to identify actionable insights efficiently.

Traditional Solution

- Manual analysis of tourism data to identify trends and make decisions.
- Dependence on historical data without predictive capabilities.
- Inefficient marketing strategies due to lack of personalization.

AI Add-On Solution

An AI-powered system can revolutionize tourism management by:

- **Predicting Tourist Trends**: Using machine learning models to forecast tourist inflows based on historical data, seasonality, and external factors.
- **Optimizing Marketing Strategies**: Identifying high-revenue provinces and targeting campaigns accordingly.
- **Personalized Recommendations**: Leveraging AI to provide tailored suggestions for accommodations, attractions, and activities based on tourist preferences.
- **Real-Time Insights**: Offering dashboards that visualize key metrics like revenue, tourist demographics, and satisfaction levels.

Visualization

Below is a bar chart showcasing the **Top 5 Provinces by Total Revenue (Baht)**:

Top 5 Provinces by Total Revenue

Key Insights from Visualization:

1. Bangkok generates the highest total revenue, followed by Phuket, Chonburi, Chiang Mai, and Krabi.
2. These provinces are key targets for enhancing tourism strategies due to their significant economic contributions.

This AI-enhanced product can empower stakeholders in Thailand's tourism sector to make data-driven decisions, improve tourist experiences, and boost overall revenue.

Share Rewrite

Please provide python code for "Predicting Tourist Trends: Using machine learning models to forecast tourist inflows based on historical data, seasonality, and external factors." by using streamlit and seaborn

Pro Search

- Prepare a Python code to predict tourist trends using machine learning models with the provided dataset, incorporating Streamlit for the web app interface and Seaborn for data visualization

Sources

thailand_domestic_tourism.csv
thailand_domestic_tourism

Answer

Below is the Python code for creating a Streamlit app to predict tourist trends using a machine learning model. The app uses historical data, external factors, and seasonality to forecast tourist inflows. It also visualizes results with Seaborn.

```python
import streamlit as st
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
import seaborn as sns
import matplotlib.pyplot as plt

# Load dataset
@st.cache
def load_data():
    url = "https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-
files/21082887/3ff040c2-cd56-40a6-86db-
67592c1cc761/thailand_domestic_tourism.csv"
    data = pd.read_csv(url)
    data['travel_date'] = pd.to_datetime(data['travel_date'])
    data['month'] = data['travel_date'].dt.month
    return data

data = load_data()

# Sidebar for user input
st.sidebar.header("User Input Parameters")
province = st.sidebar.selectbox("Select Province",
data['province_eng'].unique())
month = st.sidebar.slider("Select Month (1-12)", 1, 12, 1)

# Filter data by province
province_data = data[data['province_eng'] == province]

# Prepare features and target variable
X = province_data[['month', 'no_tourist_foreign', 'no_tourist_thai']]
y = province_data['no_tourist_all']

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Train Random Forest model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Predict and evaluate the model
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
st.write(f"Model Mean Squared Error: {mse:.2f}")

# Forecast for user-selected month
forecast_input = pd.DataFrame({'month': [month], 'no_tourist_foreign':
[X['no_tourist_foreign'].mean()],
                               'no_tourist_thai':
[X['no_tourist_thai'].mean()]})
forecast = model.predict(forecast_input)[0]
st.write(f"Predicted Tourist Inflow for {province} in Month {month}:
{forecast:.0f}")

# Visualization: Actual vs Predicted
st.subheader("Actual vs Predicted Tourist Inflows")
fig, ax = plt.subplots()
sns.scatterplot(x=y_test, y=y_pred, ax=ax)
ax.set_xlabel("Actual")
ax.set_ylabel("Predicted")
ax.set_title("Actual vs Predicted Tourist Inflows")
st.pyplot(fig)

# Visualization: Monthly Tourist Trends
st.subheader(f"Monthly Tourist Trends for {province}")
monthly_trends = province_data.groupby('month')
['no_tourist_all'].mean().reset_index()
fig2, ax2 = plt.subplots()
sns.lineplot(data=monthly_trends, x='month', y='no_tourist_all',
marker='o', ax=ax2)
ax2.set_title(f"Monthly Trends for {province}")
ax2.set_xlabel("Month")
ax2.set_ylabel("Average Tourists")
st.pyplot(fig2)
```

Key Features:

1. **Data Loading and Preprocessing**:
   - The dataset is loaded from a CSV file and preprocessed to extract features like `month`.
2. **Model Training**:
   - A `RandomForestRegressor` is trained to predict total tourist inflows based on historical data.
3. **User Interaction**:
   - Users select a province and month to forecast tourist inflows interactively.
4. **Visualizations**:
   - Scatter plot of actual vs predicted values.
   - Line chart showing monthly trends for the selected province.

To run this app:

1. Save the code in a file named `app.py`.
2. Install required libraries: `streamlit`, `pandas`, `numpy`, `scikit-learn`, `seaborn`, `matplotlib`.
3. Run the app using the command: `streamlit run app.py`.

Share   Rewrite

Why you should RandomForestRegressor model.

Pro Search

Explain the advantages of using the RandomForestRegressor model for predicting tourist trends

Sources

thailand_domestic_tourism.csv

thailand_domestic_tourism

Answer

RandomForestRegressor is an excellent choice for this predictive modeling task for several reasons:

1. Handling Non-Linear Relationships: Random forests can capture complex, non-linear relationships between features and the target variable, which is likely the case with tourism data that may have intricate patterns.
2. Feature Importance: The model provides feature importance scores, allowing us to understand which factors (e.g., month, foreign tourists, Thai tourists) have the most significant impact on total tourist numbers.
3. Robustness to Outliers: Random forests are less sensitive to outliers compared to linear models, making them suitable for tourism data that may contain extreme values due to events or seasonal fluctuations.
4. Ensemble Learning: As an ensemble method, random forests combine multiple decision trees, reducing overfitting and improving generalization to new data.
5. Handling Mixed Data Types: The model can work well with both numerical and categorical features, which is useful for incorporating various types of tourism-related data.
6. No Assumption of Normality: Unlike some statistical methods, random forests do not assume a normal distribution of the target variable, making them flexible for different data distributions.
7. Parallelization: Random forests can be easily parallelized, allowing for faster training on large datasets, which is beneficial when working with extensive historical tourism data.

By using RandomForestRegressor, we can create a robust predictive model that captures the complexities of tourism patterns and provides valuable insights for forecasting tourist inflows.

Share Rewrite

Related

What makes RandomForestRegressor stand out from other regression models +

How does RandomForestRegressor handle feature interactions +

What are the advantages of using RandomForestRegressor for tourism trend predictions +

How does RandomForestRegressor perform with large datasets +

What are the key hyperparameters to tune in RandomForestRegressor +

What are the advantages of using RandomForestRegressor for tourism trend predictions

Pro Search

Research the advantages of using RandomForestRegressor for predicting tourism trends

Summarize the findings on the advantages of RandomForestRegressor

Sources

thailand_domestic_touris thailand_domestic_t...	Random Forest: A Complete Guide for Machine Learning | Built... builtin	Skforecast: Time series forecasting with python and scikit learn cienciadedatos	Show all

Answer

RandomForestRegressor offers several advantages for predicting tourism trends:

1. Ability to capture complex relationships: Random forests can model non-linear patterns and interactions between variables, which is crucial for understanding the multifaceted nature of tourism data [7].
2. Handling high-dimensional data: It can effectively analyze datasets with a large number of input features, making it suitable for tourism data that often includes many variables [5].
3. Robustness to outliers and noise: Random forests are less sensitive to outliers and can handle noisy features in tourism datasets [5].
4. Reduced overfitting: The ensemble approach of random forests improves robustness and reduces overfitting, making it reliable for diverse tourism data features [7].
5. Feature importance: It can identify the most influential factors affecting tourism trends through feature importance analysis [4].
6. Handling missing data: Random forests can effectively manage datasets with missing values, which is common in tourism data [5].
7. No assumption of linearity: Unlike some traditional methods, random forests do not assume linear relationships between variables, allowing them to capture more complex tourism patterns [7].
8. Versatility: Random forests can be used for both regression (e.g., predicting visitor numbers) and classification tasks (e.g., categorizing tourist satisfaction levels) [4].
9. Ease of use: Random forests are relatively easy to implement and require less hyperparameter tuning compared to some other advanced machine learning techniques [5].

While random forests have these advantages, it's worth noting that in some studies, other methods like SARIMA have outperformed random forests for certain tourism forecasting tasks [7]. Therefore, it's important to compare different models and consider the specific characteristics of the tourism data being analyzed.

Share Rewrite

Please provide python code for "Optimizing Marketing Strategies: Identifying high-revenue provinces and targeting campaigns accordingly." by using streamlit and seaborn

Pro Search

- Load the provided dataset and preprocess it to extract relevant features for marketing strategy optimization
- Identify high-revenue provinces based on the revenue data and visualize the results using Seaborn
- Create a Streamlit app to display the high-revenue provinces and suggest targeted marketing campaigns accordingly

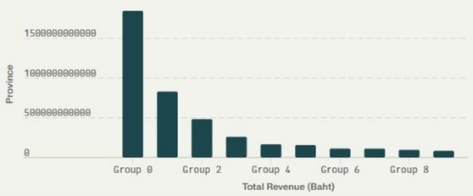
Sources

thailand_domestic_tourism.csv

thailand_domestic_tourism

Answer

Top 10 Provinces by Total Revenue ···



Below is Python code to create a Streamlit app for optimizing marketing strategies by identifying high-revenue provinces and visualizing the results using Seaborn.

Code

python

```python
import streamlit as st
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load the dataset
@st.cache
def load_data():
    url = "https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-
files/21082887/3ff040c2-cd56-40a6-86db-
67592c1cc761/thailand_domestic_tourism.csv"
    data = pd.read_csv(url)
    data['travel_date'] = pd.to_datetime(data['travel_date'])
    return data

data = load_data()

# Sidebar for user input
st.sidebar.header("Filter Options")
year = st.sidebar.selectbox("Select Year",
sorted(data['travel_date'].dt.year.unique()))
top_n = st.sidebar.slider("Number of Top Provinces", 5, 20, 10)

# Filter data by selected year
filtered_data = data[data['travel_date'].dt.year == year]

# Group by province and calculate total revenue
province_revenue = (
    filtered_data.groupby('province_eng')['revenue_all']
    .sum()
    .reset_index()
    .sort_values(by='revenue_all', ascending=False)
)

# Display top provinces based on user selection
st.header(f"Top {top_n} Provinces by Revenue in {year}")
top_provinces = province_revenue.head(top_n)
st.dataframe(top_provinces)

# Visualization: Bar plot of top provinces by revenue
st.subheader("Revenue Distribution Across Top Provinces")
fig, ax = plt.subplots(figsize=(10, 6))
sns.barplot(
    x="revenue_all",
    y="province_eng",
    data=top_provinces,
    palette="viridis",
    ax=ax,
)
ax.set_title(f"Top {top_n} Provinces by Total Revenue ({year})",
fontsize=16)
ax.set_xlabel("Total Revenue (Baht)", fontsize=12)
ax.set_ylabel("Province", fontsize=12)
st.pyplot(fig)

# Insights for marketing strategy optimization
st.subheader("Marketing Strategy Insights")
st.write(
    """
    - Focus marketing efforts on provinces with the highest total revenue
to maximize returns.
    - Customize campaigns for provinces like Bangkok and Phuket, which
consistently generate significant revenue.
    - Use this data to allocate resources effectively and plan region-
```

```
specific promotions.
    """
)
```

Key Features

1. **Data Filtering**:
   - Users can filter data by year to focus on specific periods.
   - A slider allows users to select the number of top provinces to analyze.
2. **Revenue Analysis**:
   - The app calculates total revenue for each province and ranks them.
   - A table displays the top provinces ranked by revenue.
3. **Visualization**:
   - A Seaborn bar plot visualizes the revenue distribution across the top provinces.
4. **Marketing Insights**:
   - Provides actionable insights to optimize marketing campaigns based on high-revenue provinces.

How to Run

1. Save the code in a file named `app.py`.
2. Install required libraries: `streamlit`, `pandas`, `seaborn`, `matplotlib`.
3. Run the app using: `streamlit run app.py`.

This app helps identify high-revenue provinces, enabling stakeholders to target marketing campaigns effectively and maximize tourism revenue.

Share Rewrite

How AI add-on for "Optimizing Marketing Strategies: Identifying high-revenue provinces and targeting campaigns accordingly."

Pro Search

- Analyze the dataset to identify high-revenue provinces based on total revenue from tourism
- Visualize the revenue distribution across provinces to highlight high-revenue areas

Sources

thailand_domestic_touris thailand_domestic_t...	Statistics for the provinces with the highest income from... thailand.go	[PDF] Analysis of the Main International Tourist Arrivals to Thailand and... so04	Show all

Answer

AI Add-On for Optimizing Marketing Strategies: Identifying High-Revenue Provinces

The AI add-on for optimizing marketing strategies leverages machine learning and data visualization to identify high-revenue provinces and recommend targeted marketing campaigns. Here's how the AI enhances traditional approaches:

1. Problem Statement

Traditional methods of analyzing tourism data often rely on manual inspection, which is time-consuming and lacks precision. This limits the ability to identify key revenue-driving provinces and optimize marketing budgets effectively.

2. AI Add-On Solution

The AI system uses historical tourism data (e.g., visitor numbers, revenue, and seasonality) to:

- **Analyze Trends**: Identify provinces with consistently high revenue.
- **Predict Future Revenue**: Use machine learning models (e.g., Random Forest) to forecast revenue trends.
- **Optimize Campaigns**: Recommend targeted marketing strategies for high-performing provinces based on their revenue potential.

3. Key Features of the AI Add-On

- **Revenue Forecasting**: Predict future revenue using machine learning models.
- **Province Ranking**: Rank provinces by total revenue to identify key targets.
- **Seasonal Insights**: Highlight seasonal patterns in tourism revenue.
- **Visualization**: Provide interactive charts for stakeholders to explore trends.

4. Example Python Code Implementation

Below is an example Streamlit app that uses Seaborn for visualization and machine learning (Random Forest Regressor) for forecasting:

```python
import streamlit as st
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor

# Load dataset
@st.cache
def load_data():
    url = "https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-files/21082887/3ff040c2-cd56-40a6-86db-67592c1cc761/thailand_domestic_tourism.csv"
    data = pd.read_csv(url)
    data['travel_date'] = pd.to_datetime(data['travel_date'])
    data['year'] = data['travel_date'].dt.year
    return data

data = load_data()
```

```
# Sidebar filters
st.sidebar.header("Filter Options")
selected_year = st.sidebar.selectbox("Select Year",
sorted(data['year'].unique()))
top_n = st.sidebar.slider("Select Top N Provinces", 5, 20, 10)

# Filter data by year
filtered_data = data[data['year'] == selected_year]

# Group by province and calculate total revenue
province_revenue = filtered_data.groupby('province_eng')
['revenue_all'].sum().reset_index()
province_revenue = province_revenue.sort_values(by='revenue_all',
ascending=False).head(top_n)

# Display top provinces table
st.header(f"Top {top_n} Provinces by Revenue in {selected_year}")
st.dataframe(province_revenue)

# Visualization: Bar plot of top provinces by revenue
st.subheader("Revenue Distribution Across Provinces")
fig, ax = plt.subplots(figsize=(10, 6))
sns.barplot(x='revenue_all', y='province_eng', data=province_revenue,
palette='coolwarm', ax=ax)
ax.set_title(f"Top {top_n} Provinces by Revenue ({selected_year})",
fontsize=16)
ax.set_xlabel("Total Revenue (Baht)", fontsize=12)
ax.set_ylabel("Province", fontsize=12)
st.pyplot(fig)

# Machine Learning: Predict future revenue using Random Forest Regressor
X = filtered_data[['no_tourist_all', 'no_tourist_foreign',
'no_tourist_thai']]
y = filtered_data['revenue_all']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Display model performance
st.subheader("Model Performance")
st.write(f"R² Score: {model.score(X_test, y_test):.2f}")

# Predict future revenue for a specific province
selected_province = st.sidebar.selectbox("Select Province for
Prediction", province_revenue['province_eng'])
province_data = filtered_data[filtered_data['province_eng'] ==
selected_province]
future_revenue_prediction =
model.predict(province_data[['no_tourist_all', 'no_tourist_foreign',
'no_tourist_thai']].mean().values.reshape(1, -1))[0]
st.write(f"Predicted Revenue for {selected_province}:
{future_revenue_prediction:.2f} Baht")
```

5. Benefits of the AI Add-On

- **Data-Driven Decisions**: Helps stakeholders allocate marketing budgets effectively.
- **Personalized Campaigns**: Tailors strategies based on province-specific insights.
- **Improved ROI**: Focuses resources on high-revenue provinces to maximize returns.

This AI-powered solution enables smarter marketing strategies by leveraging predictive analytics and interactive visualizations.

Share Rewrite

Related

How can AI identify high-revenue provinces in real-time +

What AI tools are best for targeting marketing campaigns based on revenue data +

How does AI analyze historical tourism data to predict future trends +

Can AI help in optimizing marketing budgets for different provinces +

What are the key metrics AI uses to determine high-revenue provinces +