
# Italinerary

Ultimate itinerary to Italy

# PROOF OF CONCEPT

# DATA COLLECTION

- **WikiVoyage**: A free web-based data-set for travel destinations and travel topics written by volunteer authors.
- It provides us data in JSON format on request with the keyword “city” as query.
- Text Preprocessing  to fetch the list of attractions.
- Creating a route by sorting them based on geographic coordinates.
- Auto-tagging  using Latent Dirichlet Allocation (LDA) algorithm.

# WIKIVOYAGE




Main page
Travel destinations
Star articles
What's Nearby?
Tourist office
Random page

Get involved
Travellers' pub
Recent changes
Community portal
Maintenance panel
Policies
Help
Interlingual lounge
Donate

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

In other projects

**Pisa** is a city in Tuscany, Italy, best known for its world-famous leaning tower. But the tower isn't the only thing to see – there are other architectural and artistic marvels in this beautiful city. The half-hour walk from the Campo dei Miracoli to the train station runs through a pedestrian street with many interesting sights, shops, and restaurants. The best way to visit Pisa is walking the streets, as the city centre is small and cosy, and enjoying the sight and the atmosphere.

## Understand [ edit ]

Pisa would not be Pisa without the university. The city is animated by the students, who organize parties, shows, and cultural events, and fill the central street of the city at night. The University of Pisa has 60,000 students in a city of 90,000 inhabitants (200,000 in the metropolitan area). You'll notice the student flair in the city once you leave the touristy Campo dei Miracoli.

### Climate [ edit ]

The climate of Pisa is something in between the one in the north and in the south of Italy.

Winters are mild: not as warm as in Catania nor as cold as in the Po Valley in the north. Temperatures almost never drop below zero. Snow is very rare.

Summers are warm, if hot, and sometimes very humid, although they are nowhere near as intolerable as in cities such as Milan, Turin and Ferrara, because they are mitigated by the wind and the sea (which is about 10 km away from the city centre). Rainfall in July and August is rare.

Spring and autumn are very pleasant, although most rainfall occurs in these seasons.

# API

https://en.wikivoyage.org/w/api.php?action=query&format=json&prop=revisions&rvprop=content&rvsection=11&titles=Pisa

{"batchcomplete":"","warnings":{"main":{"*":"Subscribe to the mediawiki-api-announce mailing list at <https://lists.wikimedia.org/mailman/listinfo/mediawiki-api-announce> for notice of API deprecations and breaking changes. Use [[Special:ApiFeatureUsage]] to see usage of deprecated features by your application.\nUnrecognized parameter: contentformat."},"revisions":{"*":"Because \"rvslots\" was not specified, a legacy format has been used for the output. This format is deprecated, and in the future the new format will always be used."}},"query":{"pages":{"27718":{"pageid":27718,"ns":0,"title":"Pisa","revisions":[{"contentformat":"text/x-wiki","contentmodel":"wikitext","*":"==See==\n[[Image:Campo dei Miracoli overview2.jpg|thumb|450px|right|The Campo dei Miracoli from above: The leaning tower is on the left, the Duomo is in the centre, the Baptistery is on the right, and part of the Camposanto is in the right foreground]]\n\nPisa is divided into 4 historical quarters. There is much more than the leaning tower in the city and several different walking itineraries are available.\n\n===Piazza dei Miracoli===\nThe '''[http://www.opapisa.it/en/miracles-square/the-square.html Piazza dei Miracoli]''' or ''Field of Miracles'' is to the North of central Pisa. It is an [[UNESCO World Heritage List|UNESCO World Heritage site]] and contains the city's most famous sights:\n*{{see\n| name=Torre Pendente | alt=Leaning Tower | url=http://www.opapisa.it/ | email=info@opapisa.it\n| address=Piazza Arcivescovado 1 | lat=43.72301 | long=10.39663 | directions=\n| phone=+39 50 835011 | tollfree= | fax=\n| hours=Daily 09:00-20:00 | price=\u20ac18\n| lastedit=2019-08-30\n| wikipedia=Leaning Tower of Pisa | image=The Leaning Tower of Pisa SB.jpeg | wikidata=Q39054\n| content=The structure was conceived as the cathedral's bell tower. Construction began in 1173 and the tower started leaning soon afterwards due to subsidence of the ground underneath its base. A project to keep the tower from leaning more and tipping over finally reached a successful conclusion in 2001, and the tower is again open to those wishing to climb it. Climbing the tower requires a reservation-based ticket for \u20ac18. Tickets can be bought for the tower on the day, for a specific entry time. This could be 45 min-2 hr after the purchase time, but there is a lot to see while you wait. It is better if you [http://boxoffice.opapisa.it/Turisti/ buy tickets online] for \u20ac18 well in advance (up to 20 days). The tickets are non-exchangeable, effectively non-refundable, and only good for the tower, so they're a bit of a risk to purchase in advance. Make the effort to climb, though, and you'll be rewarded by the view. The famous Pisa leaning tower is not the only one, due of the marshy land that they are built on, there are other 2 towers in Pisa: the Bell Tower of {{marker|type=see|name=San Nicola Church|wikidata=Q3671449}}, near the banks of Arno and the Bell Tower of {{marker|type=see|name=San Michele of Scalzi Church|wikidata=Q3502492}}. For safety reasons, children who will not have turned 8 by the end of this year are not permitted to enter. Under-18s must be accompanied by an adult. ID may be requested to certify the age.\n}}\n* {{see\n| name=Duomo di Pisa | alt=Pisa Cathedral | url=http://www.opapisa.it/en/miracles-square/cathedral.html {{dead link|October 2017|August 2018}} | email=\n| address= | lat=43.723272 | long=10.395619 | directions=\n| phone= | tollfree= | fax=\n| hours= | price=Free (but entry passes, up to 2 per person, must be obtained beforehand in a nearby ticket office)\n| wikipedia=Pisa Cathedral | image=Campo dei Miracoli overview2.jpg | wikidata=Q1754247\n|lastedit=2018-08-19| content=A splendid cathedral, containing artwork by Giambologna, Della Robbia, and other major artists. Fine Romanesque style with double aisles and a cupola, a huge apse mosaic partly by Cimabue, and a fine pulpit by Giovanni Pisano in late Gothic/early Renaissance style.\n}}\n* {{see\n| name=Battistero di San Giovanni | alt=Baptistry of St. John | url=https://www.opapisa.it/en/square-of-miracles/baptistery/ | email=\n| address= | lat=43.723333 | long=10.393889 | directions=\n| phone= | tollfree= | fax=\n| hours= | price=Single ticket \u20ac5. A combined ticket with two museums is \u20ac7 and three museums is \u20ac8. It can be combined with the Monumental Cemetery and the Sinopie Museum\n| wikipedia=Pisa Baptistery | image=PisaBaptistry20020323 rectilinear.jpg | wikidata=Q1278477\n| lastedit=2019-08-30\n| content=Large round Romanesque dome with many sculptured decorations and a fine view up top; climb this if you want a great view with the Leaning Tower visible in your photos. Arabic-style pavement, pulpit by Nicola Pisano (father of Giovanni), and fine octagonal font. At regular intervals, the ticket-checker-guard at the entrance comes into the baptistery and gives an audio-treat of echo-effect. The guard shouts out few sounds which when echoed sound like pure beautiful music. You can also cast your inhibitions to the wind, stand by the wall, and sing long notes that turn into chords by yourself, as the echoes go round and round the dome of the building.\n}}\n* {{see\n| name=Camposanto Monumentale | alt=Monumental Cemetery | url=https://www.opapisa.it/en/square-of-miracles/camposanto/ | email=\n| address= | lat=43.723 | long=10.395 | directions=\n| phone= | tollfree= | fax=\n| hours= | price=Single ticket \u20ac5. A combined ticket with two museums is \u20ac7 and three museums is \u20ac8. It can be combined with the Baptistery and the Sinopie Museum. Free entrance to the cemetery on 1 and 2 Nov\n| wikipedia=Camposanto Monumentale | image=Pisa.Camposanto01.jpg | wikidata=Q1031614\n| lastedit=2019-08-30\n| content=A huge cemetery building with lots of interesting art, including a collection of ancient Roman sarcophagi and splendid medieval frescoes by the \"Master of the Triumph of Death\". There is also a 19th century statue of the famous mathematician Leonardo Fibonacci, a native of the city.\n}}\n* {{see\n| name=Museo dell'Opera del Duomo | alt= | url=https://www.opapisa.it/en/square-of-miracles/opera-del-duomo-museum/ | email=\n| address= | lat=43.7225 | long=10.3969 | directions=\n| phone= | tollfree= | fax=\n| hours=Temporarily closed for restoration | price=\n| image=Opae logo.jpg | wikidata=Q3328407\n| lastedit=2019-08-30\n| content=Has sculptures and paintings that used to be preserved in the Cathedral and the cemetery. Some of the more unusual are bronze griffins from Syria captured by the Crusaders. You can also capture nice photos from the Tower and the Duomo from its balcony.\n}}\n* {{see\n| name=Museo delle Sinopie | alt= | url=http://www.opapisa.it/it/organizza-la-tua-visita/informazioni-pratiche/orari.html | email=\n| address= | lat=43.7223 | long=10.3947 | directions=\n| phone= | tollfree= | fax=\n| hours= | price=Single ticket \u20ac5. A combined ticket with two museums is \u20ac7 and three museums is \u20ac8. It can be combined with the Baptistery and the Monumental Cemetery\n| image=Interior del Museu delle Sinopie de Pisa.JPG | wikidata=Q3328416\n| lastedit=2019-08-30\n| content=Skipped over by many visitors, this museum is a treat for art lovers. After World War II many of the surviving murals and pieces of murals from Pisa's Campo Santo were detached from the walls to try to preserve them. It was unexpectedly discovered that the artist

# CLEANING DATA

```
import requests
import json
import re
import codecs
import pprint

#Fetching data
URL =
"https://en.wikivoyage.org/w/api.php?action=query&format=json&prop=revisions&rvp
rop=content&rvsection=11&titles=Pisa"
r = requests.get(url = URL)

#Fetching data related to our query
data = r.json()
querydata=data["query"]["pages"]

#converting to string for parsing
detaileddata="".join(str(items) for items in querydata.values())

#parsing to get the attraction list
datalist=re.findall(r"name=(.+?)\|",detaileddata)
```

```
#cleaning
finallist=[]
for i in datalist:
    finallist.append(codecs.decode(i, 'unicode_escape'))

#fetching latitude longitude
j=0
latlonglist=[]
for i in finallist:
    idict={}
    placename=i.split("\"")
    regexp=placename[-1]+"(.+?)directions"
    try:
        latlongdata=re.search(regexp,detaileddata).group(1)
        latdata=re.search(r"lat=(.+?)\|",latlongdata).group(1)
        longdata=re.search(r"long=(.+?)\|",latlongdata).group(1)
        idict["name"]=i
        idict["lat"]=float(latdata)
        idict["lon"]=float(longdata)
        latlonglist.append(idict)
    except:
            pass
pp = pprint.PrettyPrinter(depth=3)
pp.pprint("List:",latlonglist)
```

# OUTPUT


[{'lat': 43.72301, 'lon': 10.39663, 'name': 'Torre Pendente '},
{'lat': 43.723272, 'lon': 10.395619, 'name': 'San Nicola Church'},
{'lat': 43.723272, 'lon': 10.395619, 'name': 'San Michele of Scalzi Church'},
{'lat': 43.723272, 'lon': 10.395619, 'name': 'Duomo di Pisa '},
{'lat': 43.723333, 'lon': 10.393889, 'name': 'Battistero di San Giovanni '},
{'lat': 43.723, 'lon': 10.395, 'name': 'Camposanto Monumentale '},
{'lat': 43.7225, 'lon': 10.3969, 'name': "Museo dell'Opera del Duomo "},
{'lat': 43.7223, 'lon': 10.3947, 'name': 'Museo delle Sinopie '},
{'lat': 43.719444, 'lon': 10.4, 'name': 'Piazza dei Cavalieri '},
{'lat': 43.719611, 'lon': 10.400225, 'name': 'Palazzo della Carovana '},
{'lat': 43.719611, 'lon': 10.400225, 'name': "Palazzo dell'Orologio "},
{'lat': 43.719325, 'lon': 10.400953, 'name': 'Chiesa di Santo Stefano '},
{'lat': 43.714533, 'lon': 10.407614, 'name': 'Museo Nazionale di San Matteo '},
{'lat': 43.71644,'lon': 10.40223, 'name': "Piazza Garibaldi & Piazza XX Settembre "},
{'lat': 43.714468, 'lon': 10.403558, 'name': 'Santo Sepolcro '},
{'lat': 43.715278, 'lon': 10.396389, 'name': 'Santa Maria della Spina '},
{'lat': 43.711031, 'lon': 10.406767, 'name': 'Giardino Scotto '},
{'lat': 43.7143, 'lon': 10.3896, 'name': 'La Cittadella '},
{'lat': 43.719722, 'lon': 10.395833, 'name': 'University Botanical Garden '},
{'lat': 43.716789,'lon': 10.390936, 'name': 'Museum of Computing Instruments '},
{'lat': 43.711111, 'lon': 10.3975, 'name': 'Tuttomondo '}]


# CREATING ROUTE

```
from math import sin, cos, sqrt, atan2, radians
def distance(lat1, lon1, lat2, lon2):
    R = 6373.0
    lat1 = radians(lat1)
    lon1 = radians(lon1)
    lat2 = radians(lat2)
    lon2 = radians(lon2)
    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = sin(dlat / 2)**2 + cos(lat1) * cos(lat2) * sin(dlon / 2)**2
    c = 2 * atan2(sqrt(a), sqrt(1 - a))
    return R * c
def closest(data, v):
    if len(data)!=0:
        return min(data, key=lambda p: distance(v['lat'],v['lon'],p['lat'],p['lon']))
o = {'name': 'Torre Pendente ', 'lat': 43.72301, 'lon': 10.39663}
v = closest(latlonglist, o)
i=1
tempList= latlonglist
while len(tempList)!=0:
    tempList = [x for x in tempList if not x == v]
    print(i,".",v['name'])
    v = closest(tempList, o)
    i = i+1
```

## OUTPUT

1 . Torre Pendente
2 . Museo dell'Opera del Duomo
3 . San Nicola Church
4 . San Michele of Scalzi Church
5 . Duomo di Pisa
6 . Camposanto Monumentale
7 . Museo delle Sinopie
8 . Battistero di San Giovanni
9 . University Botanical Garden
10 . Palazzo della Carovana
11 . Palazzo dell'Orologio
12 . Piazza dei Cavalieri
13 . Chiesa di Santo Stefano
14 . Museum of Computing Instruments
15 . Piazza Garibaldi "'and"' Piazza XX Settembre
16 . Santa Maria della Spina
17 . Santo Sepolcro
18 . La Cittadella
19 . Museo Nazionale di San Matteo
20 . Tuttomondo
21 . Giardino Scotto

# AUTO-TAGGING USING LDA

**TEXT**

Pisa is a city in Tuscany, Italy, best known for its world-famous leaning tower with other beautiful ,architecture and artistic marvels.

**TOPICS**

artistic

architecture

COMPARE

**PREDEFINED TAGS**

artistic

beach

architecture

romantic

historic

…..

# INPUT

```
from sklearn.feature_extraction.text import CountVectorizer
```

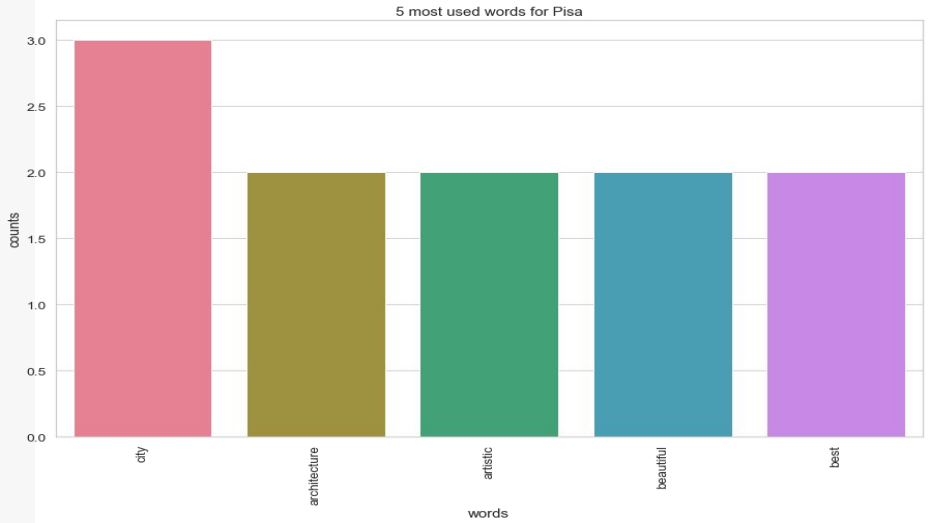
```
import pandas as pd
import numpy as np
import requests


import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('whitegrid')
%matplotlib inline
```

```
url = "https://en.wikivoyage.org/w/api.php?format=json&action=query&prop=extracts&exintro&explaintext&redirects=1&titles=Pisa"

f = requests.get(url)
print(f.text)
```


```
{"batchcomplete":"","query":{"pages":{"27718":{"pageid":27718,"ns":0,"title":"Pisa","extract":"Pisa is a city in Tuscany, Ital
y, best known for its world-famous leaning tower. But the tower isn't the only thing to see \u2013 there are other architectura
l and artistic marvels in this beautiful city. The half-hour walk from the Campo dei Miracoli to the train station runs through
a pedestrian street with many interesting sights, shops, and restaurants. The best way to visit Pisa is walking the streets, as
the city centre is small and cosy, and enjoying the sight and the atmosphere."}}}}
```

# COMMON WORDS

```
def plot_5_most_common_words(count_data, count_vectorizer):
    import matplotlib.pyplot as plt
    words = count_vectorizer.get_feature_names()
    total_counts = np.zeros(len(words))
    for t in count_data:
        total_counts+=t.toarray()[0]


    count_dict = (zip(words, total_counts))
    count_dict = sorted(count_dict, key=lambda x:x[1], reverse=True)[0:5]

    words = [w[0] for w in count_dict]
    counts = [w[1] for w in count_dict]
    print (count_dict)
    x_pos = np.arange(len(words))


    plt.figure(2, figsize=(15, 15/1.6180))
    plt.subplot(title='5 most used words for Pisa')
    sns.set_context("notebook", font_scale=1.25, rc={"lines.linewidth": 2.5})
    sns.barplot(x_pos, counts, palette='husl')
    plt.xticks(x_pos, words, rotation=90)
    plt.xlabel('words')
    plt.ylabel('counts')
    plt.show()


count_vectorizer = CountVectorizer(stop_words='english')
count_data = count_vectorizer.fit_transform(data)

plot_5_most_common_words(count_data, count_vectorizer)
```



# LDA MODEL

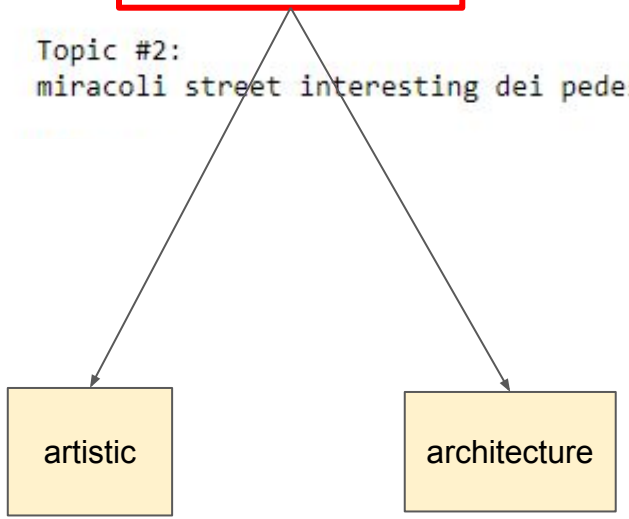
```
import warnings
warnings.simplefilter("ignore", DeprecationWarning)
```

```
# Load the LDA model from sk-learn
from sklearn.decomposition import LatentDirichletAllocation as LDA
```

```
# Helper function
def print_topics(model, count_vectorizer, n_top_words):
    words = count_vectorizer.get_feature_names()

    for topic_idx, topic in enumerate(model.components_):

        print("\nTopic #%d:" % topic_idx)
        print(" ".join([words[i]
                        for i in topic.argsort()[:-n_top_words - 1:-1]]))
```

```
# Tweak the two parameters below (use int values below 15)
number_topics = 5
number_words = 5


# Create and fit the LDA model
lda = LDA(n_components=number_topics)
lda.fit(count_data)
```

```
Topic #0:
pisa best city streets sight

Topic #1:
city architecture artistic beautiful tower

Topic #2:
miracoli street interesting dei pedestrian
```



# FINAL APPLICATION

Scale it for all cities in Italy

Fetch Restaurant and transport options

Create itinerary from the database based on user preferences