

SKRIPSI

**ANALISIS DAN VISUALISASI DATA INDUSTRI PARIWISATA
PROVINSI JAWA BARAT**



Syahdan Riyantyo Putro

NPM: 6181801043

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2023**

UNDERGRADUATE THESIS

**DATA ANALYSIS AND VISUALIZATION OF THE WEST JAVA
PROVINCE TOURISM INDUSTRY**



Syahdan Riyantyo Putro

NPM: 6181801043

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2023**

LEMBAR PENGESAHAN

ANALISIS DAN VISUALISASI DATA INDUSTRI PARIWISATA PROVINSI JAWA BARAT

Syahdan Riyantyo Putro

NPM: 6181801043

Bandung, 4 Juli 2023

Menyetujui,

Pembimbing

Digitally signed
by Cecilia Esti
Nugraheni

Dr.rer.nat. Cecilia Esti Nugraheni

Ketua Tim Penguji

Digitally signed
by Pascal

Alfadian Nugroho

Pascal Alfadian, Nugroho, M.Comp.

Anggota Tim Penguji

Digitally signed
by Rosa de Lima
E. Padmowati

Rosa De Lima, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed
by Mariskha Tri
Adithia

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS DAN VISUALISASI DATA INDUSTRI PARIWISATA PROVINSI JAWA BARAT

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 4 Juli 2023



Syahdan Riyantyo Putro
NPM: 6181801043

ABSTRAK

Industri pariwisata adalah kumpulan usaha pariwisata yang menghasilkan barang dan/atau jasa bagi pemenuhan kebutuhan wisatawan dalam berpariwisata. Industri pariwisata meliputi akomodasi, kegiatan layanan makanan dan minuman, angkutan penumpang, agen perjalanan wisata, kegiatan budaya, dan kegiatan olahraga dan hiburan. Pada tahun 2020 Indonesia dilanda wabah penyakit pernapasan yaitu COVID-19 yang membuat Pemerintah Indonesia membuat kebijakan untuk mengatasi penyebaran virus COVID-19 dimana salah satu kebijakan itu adalah Pembatasan Sosial Berskala Besar (PSBB). Akibatnya masyarakat tidak dapat melakukan perjalanan wisata sehingga industri pariwisata terkena dampak yang sangat besar. Penelitian ini memanfaatkan teknik-teknik *data science* untuk mempelajari data-data lampau untuk memperoleh *insight* untuk menentukan strategi yang tepat demi membangkitkan industri pariwisata. Penelitian ini menggunakan bahasa pemrograman R untuk memperoleh *insight* dari data lampau dan dibuat sebuah *website dashboard* dengan menggunakan *RShiny*.

Data science merupakan cabang ilmu yang mengkombinasikan koding, statistika dan bisnis untuk memperoleh *insight*. Proses *data science* yang dilakukan dalam penelitian ini adalah pengumpulan data yang berkaitan dengan industri pariwisata. Penyiapan data untuk mengatasi *missing value* dalam data. Visualisasi data untuk melihat pola-pola menarik dari data. Korelasi adalah nilai koefisien untuk mengukur korelasi/hubungan antara dua variabel numerik.

Pada penelitian ini terdapat dua *dataset* yang dianalisis. Pertama merupakan data terkait persentase tingkat hunian kamar hotel. Kedua merupakan gabungan dari beberapa *dataset* yang meliputi jumlah kamar hotel bintang 1 sampai dengan bintang 5 dan non bintang, jumlah kamar hotel bintang 1 sampai dengan bintang 5, jumlah restoran, jumlah angkutan wisata, jumlah objek wisata alam, budaya, dan minat khusus, jumlah wisata nusantara dan jumlah wisatawan mancanegara.

Visualisasi yang digunakan dalam penelitian ini adalah histogram yang berguna untuk melihat distribusi dari data untuk menentukan *missing value* diisi dengan *mean*, *median*, atau dengan nilai observasi sebelumnya yang tidak hilang. *Trend* dari *dataset* persentase hunian divisualisasikan dengan *line plot* untuk melihat bulan apa saja persentase hunian meningkat. Analisis korelasi dilakukan terhadap *dataset* gabungan untuk melihat variabel yang mempengaruhi peningkatan jumlah wisatawan nusantara dan mancanegara. *Scatter plot* dibuat untuk masing-masing variabel yang berpengaruh terhadap wisatawan nusantara dan mancanegara untuk memvalidkan hasil dari nilai korelasi.

Berdasarkan hasil analisis dan visualisasi yang dilakukan, dapat disimpulkan bahwa persentase hunian kamar hotel tertinggi paling sering terdapat pada bulan November dan Desember. Variabel-variabel yang memiliki hubungan positif dengan wisatawan nusantara adalah hotel non-bintang, jumlah hotel keseluruhan, jumlah restoran, dan jumlah objek wisata. Untuk wisatawan mancanegara tidak terdapat variabel-variabel yang berhubungan positif.

Kata-kata kunci: Industri pariwisata, Data science, *Insight*, Visualisasi data

ABSTRACT

Tourism industry is a collection of tourism businesses that produce goods and/or services to meet the needs of tourists in their tourism activities. The tourism industry includes accommodations, food and beverage services, passenger transportation, travel agencies, cultural activities, and sports and entertainment activities. In 2020, Indonesia was hit by a respiratory disease outbreak, namely COVID-19, which led the Indonesian Government to implement policies to control the spread of the virus, including Large-Scale Social Restrictions (PSBB). As a result, people were unable to travel for tourism purposes, and the tourism industry was severely affected. This research utilizes data science techniques to study past data in order to gain insights and determine appropriate strategies to revive the tourism industry. The research utilizes the R programming language to extract insights from past data and creates a website dashboard using RShiny.

Data science is a field that combines coding, statistics, and business knowledge to obtain insights. The data science process conducted in this research involves data collection related to the tourism industry, data preparation to handle missing values, data visualization to identify interesting patterns, and correlation analysis to measure the correlation/relationship between two numerical variables.

In this research, two datasets are analyzed. The first dataset relates to the occupancy rate of hotel rooms. The second dataset is a combination of several datasets, including the number of hotel rooms for 1 to 5-star hotels and non-star hotels, the number of 1 to 5-star hotel rooms, the number of restaurants, the number of tourist transportation, the number of natural, cultural, and special interest tourist attractions, and the number of domestic and international tourists.

The visualizations used in this research include histograms, which are useful for observing data distribution and determining how to fill missing values with mean, median, or previous observation values. The occupancy rate dataset is visualized using a line plot to identify the months with the highest occupancy rates. Correlation analysis is performed on the combined dataset to identify variables that influence the increase in the number of domestic and international tourists. Scatter plots are created for each variable that affects domestic and international tourists to validate the correlation analysis results.

Based on the analysis and visualizations conducted, it can be concluded that the highest occupancy rates are most commonly observed in November and December. Variables that have a positive correlation with domestic tourists are non-star hotels, the total number of hotels, the number of restaurants, and the number of tourist attractions. There are no variables positively correlated with international tourists.

Keywords: Tourism industry, Data science, Insight, Data visualization

*Saya persembahkan skripsi ini kepada ibu, almarhum ayah, dan
Alvarestha...*

KATA PENGANTAR

Dengan mengucapkan Alhamdulillah segala puji dan syukur penulis panjatkan atas kehadiran Allah SWT, karena berkat rahmat dan hidayah-Nya penyusunan skripsi yang berjudul “Analisis Dan Visualisasi Data Industri Pariwisata Provinsi Jawa Barat” ini dapat diselesaikan guna memenuhi salah satu persyaratan dalam menyelesaikan pendidikan pada Jurusan Informatika Fakultas Teknologi Informasi Dan Sains Universitas Katolik Parahyangan Bandung. Penulis menyadari bahwa penyusunan skripsi ini tidak akan berhasil tanpa dukungan doa dari berbagai pihak, oleh karena itu penulis ingin mengucapkan terima kasih kepada:

- Ibu penulis yang telah mendoakan, mendukung dan memenuhi kebutuhan penulis selama proses penyusunan skripsi.
- Alvarestha yang selalu mendoakan dan memberikan dukungan selama proses penyusunan skripsi
- Ibu Dr.Rer.Nat. Cecilia Esti Nugraheni, ST, MT. yang telah memberikan bimbingan dan arahan selama proses penyusunan skripsi.
- Sahabat semasa kuliah Cevas, Widad, Soetomo, Fachri, Tjahyadi, dan David.
- Teman-teman lain yang tidak dapat penulis sebutkan satu per satu yang sudah memberikan dukungan dalam mengerjakan skripsi ini.

Penulis berharap semoga skripsi ini dapat berguna bagi segenap pihak yang berkepentingan. Akhir kata, penulis memohon maaf apabila terdapat kekurangan dalam hasil penyusunan skripsi ini.

Bandung, Juli 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxv
DAFTAR KODE PROGRAM	xxvii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	4
1.4 Batasan Masalah	4
1.5 Metodologi	4
1.6 Sistematika Pembahasan	4
2 LANDASAN TEORI	7
2.1 Industri Pariwisata	7
2.2 <i>Dataset</i> Industri Pariwisata	7
2.2.1 Badan Pusat Statistik Jawa Barat	7
2.2.2 <i>Open Data</i> Jabar	8
2.3 <i>Preprocessing Data</i>	8
2.4 Distribusi Normal	9
2.5 Distribusi Miring	9
2.6 Uji Shapiro-Wilk	10
2.7 <i>Mean</i>	11
2.8 <i>Median</i>	11
2.9 Visualisasi Data	11
2.9.1 Line Chart	12
2.9.2 Scatter Plot	12
2.9.3 Histogram	13
2.10 Korelasi, Matriks Korelasi, dan <i>Correlation Plot</i>	13
3 <i>Preprocessing Data</i> DAN EKSPLORASI	17
3.1 Analisis Masalah	17
3.2 Pengumpulan Data	17
3.3 Penyiapan Data	18
3.3.1 <i>Dataset Open Data</i> Jabar	18
3.3.2 <i>Dataset</i> Persentase Hunian	26
3.4 Eksplorasi <i>Dataset</i>	27
3.4.1 Eksplorasi Jumlah Wisatawan	28

3.4.2	Hasil Eksplorasi Jumlah Wisatawan	33
3.4.3	Eksplorasi Jumlah Hotel	33
3.4.4	Hasil Eksplorasi Jumlah Hotel	60
3.4.5	Eksplorasi Jumlah Restoran, Rumah Makan, dan <i>Cafe</i>	60
3.4.6	Hasil Eksplorasi Jumlah Restoran, Rumah Makan, dan <i>Cafe</i>	63
4	ANALISIS DAN PENCARIAN <i>Insight</i>	65
4.1	Deskripsi <i>Dataset</i>	65
4.2	Analisis <i>Dataset</i> dan Penggalian <i>Insight</i>	69
4.2.1	Trend Persentase Hunian	70
4.2.2	Hasil Persentase Hunian	71
4.2.3	Analisis Variabel Independen	71
4.2.4	Hasil Variabel Independen	92
5	PERANGKAT LUNAK	95
5.1	Rancangan <i>Website Dashboard</i>	95
5.2	Hasil <i>Website Dashboard</i>	96
6	KESIMPULAN DAN SARAN	101
6.1	Kesimpulan	101
6.2	Saran	102
	DAFTAR REFERENSI	103
	A KODE PROGRAM	105
	B HASIL EKSPERIMEN	143
	C DATASET	171

DAFTAR GAMBAR

1.1 Tahapan <i>Data Science</i> [1]	2
1.2 Contoh-Contoh Visualisasi Data[2]	2
1.3 <i>Clustered Barplot</i> Jumlah Wisatawan Mancanegara	3
2.1 Gambar Histogram Berdistribusi Normal[2]	9
2.2 Bentuk Dari Distribusi Miring[3]	10
2.3 Gambar <i>Line Chart</i> [4]	12
2.4 Contoh <i>Scatter Plot</i> [2]	12
2.5 Gambar Histogram[2]	13
2.6 Contoh <i>Scatter plot</i> beserta korelasinya. Baris pertama menunjukkan hubungan positif dan baris kedua menunjukkan hubungan negatif.[5]	14
2.7 Bentuk Dari <i>Correlation Plot</i> [6][7]	15
3.1 Kode Untuk Membuat <i>Dataset</i> Kecil	21
3.2 Kode Untuk Mendeteksi <i>Missing Value</i>	21
3.3 Kode untuk melakukan uji Shapiro-Wilk	21
3.4 Hasil uji Shapiro-Wilk Untuk Data Mancanegara Di Kabupaten Bandung	22
3.5 Kode Untuk Membuat Histogram	22
3.6 Hasil Histogram Dari Data Mancanegara Di Kabupaten Bandung	22
3.7 Kode Untuk Mengisi <i>Missing Value</i> Dengan Nilai <i>Median</i>	23
3.8 Hasil uji Shapiro-Wilk untuk data hotel non-bintang di Kabupaten Bandung	23
3.9 Hasil histogram dari data hotel non-bintang di Kabupaten Bandung	24
3.10 Hasil uji Shapiro-Wilk untuk data bus di Kabupaten Bandung. Dapat dilihat bahwa <i>error</i> yang terjadi dikarenakan seluruh variabel x bernilai identik.	25
3.11 Hasil histogram dari data yang bernilai identik	25
3.12 Kode untuk mengisi <i>missing value</i> dengan nilai sebelumnya yang tidak hilang	25
3.13 Trend Jumlah Wisatawan Nusantara (1)	28
3.14 Trend Jumlah Wisatawan Nusantara (2)	28
3.15 Trend Jumlah Wisatawan Nusantara (3)	29
3.16 Trend Jumlah Wisatawan Nusantara (4)	29
3.17 Trend Jumlah Wisatawan Nusantara (5)	29
3.18 Trend Jumlah Wisatawan Nusantara (6)	30
3.19 Trend Jumlah Wisatawan Nusantara (7)	30
3.20 Trend Jumlah Wisatawan Mancanegara (1)	31
3.21 Trend Jumlah Wisatawan Mancanegara (2)	31
3.22 Trend Jumlah Wisatawan Mancanegara (3)	31
3.23 Trend Jumlah Wisatawan Mancanegara (4)	32
3.24 Trend Jumlah Wisatawan Mancanegara (5)	32
3.25 Trend Jumlah Wisatawan Mancanegara (6)	32
3.26 Trend Jumlah Wisatawan Mancanegara (7)	33
3.27 Trend Jumlah Hotel Berbintang Di Kabupaten Bandung	34
3.28 Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Bandung	34
3.29 Trend Jumlah Hotel Berbintang Di Kabupaten Bandung Barat	35

3.30	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Bandung Barat	35
3.31	Trend Jumlah Hotel Berbintang Di Kabupaten Bekasi	36
3.32	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Bekasi	36
3.33	Trend Jumlah Hotel Berbintang Di Kabupaten Bogor	37
3.34	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Bogor	37
3.35	Trend Jumlah Hotel Berbintang Di Kabupaten Ciamis	38
3.36	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Ciamis	38
3.37	Trend Jumlah Hotel Berbintang Di Kabupaten Cianjur	39
3.38	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Cianjur	39
3.39	Trend Jumlah Hotel Berbintang Di Kabupaten Cirebon	40
3.40	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Cirebon	40
3.41	Trend Jumlah Hotel Berbintang Di Kabupaten Garut	41
3.42	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Garut	41
3.43	Trend Jumlah Hotel Berbintang Di Kabupaten Indramayu	42
3.44	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Indramayu	42
3.45	Trend Jumlah Hotel Berbintang Di Kabupaten Karawang	43
3.46	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Karawang	43
3.47	Trend Jumlah Hotel Berbintang Di Kabupaten Kuningan	44
3.48	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Kuningan	44
3.49	Trend Jumlah Hotel Berbintang Di Kabupaten Majalengka	45
3.50	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Majalengka	45
3.51	Trend Jumlah Hotel Berbintang Di Kabupaten Pangandaran	46
3.52	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Pangandaran	46
3.53	Trend Jumlah Hotel Berbintang Di Kabupaten Purwakarta	47
3.54	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Purwakarta	47
3.55	Trend Jumlah Hotel Berbintang Di Kabupaten Subang	48
3.56	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Subang	48
3.57	Trend Jumlah Hotel Berbintang Di Kabupaten Sukabumi	49
3.58	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Sukabumi	49
3.59	Trend Jumlah Hotel Berbintang Di Kabupaten Sumedang	50
3.60	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Sumedang	50
3.61	Trend Jumlah Hotel Berbintang Di Kabupaten Tasikmalaya	51
3.62	Trend Jumlah Hotel Tidak Berbintang Di Kabupaten Tasikmalaya	51
3.63	Trend Jumlah Hotel Berbintang Di Kota Bandung	52
3.64	Trend Jumlah Hotel Tidak Berbintang Di Kota Bandung	52
3.65	Trend Jumlah Hotel Berbintang Di Kota Banjar	53
3.66	Trend Jumlah Hotel Tidak Berbintang Di Kota Banjar	53
3.67	Trend Jumlah Hotel Berbintang Di Kota Bekasi	54
3.68	Trend Jumlah Hotel Tidak Berbintang Di Kota Bekasi	54
3.69	Trend Jumlah Hotel Berbintang Di Kota Bogor	55
3.70	Trend Jumlah Hotel Tidak Berbintang Di Kota Bogor	55
3.71	Trend Jumlah Hotel Berbintang Di Kota Cimahi	56
3.72	Trend Jumlah Hotel Tidak Berbintang Di Kota Cimahi	56
3.73	Trend Jumlah Hotel Berbintang Di Kota Cirebon	57
3.74	Trend Jumlah Hotel Tidak Berbintang Di Kota Cirebon	57
3.75	Trend Jumlah Hotel Berbintang Di Kota Depok	58
3.76	Trend Jumlah Hotel Tidak Berbintang Di Kota Depok	58
3.77	Trend Jumlah Hotel Berbintang Di Kota Sukabumi	59
3.78	Trend Jumlah Hotel Tidak Berbintang Di Kota Sukabumi	59
3.79	Trend Jumlah Hotel Berbintang Di Kota Tasikmalaya	60
3.80	Trend Jumlah Hotel Tidak Berbintang Di Kota Tasikmalaya	60

3.81	Trend Jumlah Resto, Rumah Makan, dan <i>Cafe</i> (1)	61
3.82	Trend Jumlah Resto, Rumah Makan, dan <i>Cafe</i> (2)	61
3.83	Trend Jumlah Resto, Rumah Makan, dan <i>Cafe</i> (3)	62
3.84	Trend Jumlah Resto, Rumah Makan, dan <i>Cafe</i> (4)	62
3.85	Trend Jumlah Resto, Rumah Makan, dan <i>Cafe</i> (5)	62
3.86	Trend Jumlah Resto, Rumah Makan, dan <i>Cafe</i> (6)	63
3.87	Trend Jumlah Resto, Rumah Makan, dan <i>Cafe</i> (7)	63
4.1	Persentase Hunian Tahun 2010 Sampai 2021	70
4.2	Hasil Nilai Korelasi Untuk <i>Dataset</i> Gabungan	72
4.3	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Hotel Bintang 3	73
4.4	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Kamar Hotel Bintang 1	74
4.5	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Hotel Non-Bintang	75
4.6	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Hotel	76
4.7	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Restoran, Rumah Makan, dan <i>Cafe</i>	77
4.8	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Angkutan Wisata	78
4.9	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Objek Wisata Alam	79
4.10	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Objek Wisata Budaya	80
4.11	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Objek Wisata Minat Khusus	81
4.12	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Objek Wisata	82
4.13	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Hotel Bintang 3	83
4.14	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Hotel Non-Bintang	84
4.15	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Total Hotel	85
4.16	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Kamar Hotel Bintang 1	86
4.17	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Objek Wisata Alam	87
4.18	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Kamar Hotel Bintang 3	88
4.19	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Kamar Hotel Bintang 4	89
4.20	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Kamar Hotel	90
4.21	Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Restoran, Rumah Makan, dan <i>Cafe</i>	91
4.22	Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Total Angkutan Wisata	92
5.1	Halaman Utama Dari <i>Website Dashboard</i>	95
5.2	Halaman Utama Dari <i>Website Dashboard</i> Setelah Menekan <i>icon hamburger</i>	96
5.4	Hasil Halaman Awal Dari <i>Website Dashboard</i>	97
5.5	Hasil Halaman Awal Dari <i>Website Dashboard</i> Setelah Menekan <i>Icon Hamburger</i>	97
5.6	Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi Yang Telah Dibuat	98
5.7	Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi <i>Trend</i> Dari Data	98

5.8	Hasil Halaman Dari <i>Website Dashboard</i> Yang Memiliki Submenu Pada Menu Di <i>Sidebar</i>	99
5.9	Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi Distribusi Data	99
5.10	Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi <i>Correlation Plot</i>	100
5.11	Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi <i>Scatterplot</i>	100
B.1	Hasil Trend Persentase Hunian Tahun 2010–2012	143
B.2	Hasil Trend Persentase Hunian Tahun 2013–2015	143
B.3	Hasil Trend Persentase Hunian Tahun 2016–2018	144
B.4	Hasil Trend Persentase Hunian Tahun 2019–2021	144
B.5	Hasil Nilai Korelasi Untuk Dataset Gabungan	145
B.6	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Hotel Bintang 3	146
B.7	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Kamar Hotel Bintang 1	147
B.8	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Hotel Non-Bintang	148
B.9	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Hotel	149
B.10	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Restoran, Rumah Makan, dan <i>Cafe</i>	150
B.11	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Angkutan Wisata	151
B.12	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Objek Wisata Alam	152
B.13	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Objek Wisata Budaya	153
B.14	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Objek Wisata Minat Khusus	154
B.15	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Objek Wisata	155
B.16	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Hotel Bintang 3	156
B.17	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Hotel Non-Bintang	157
B.18	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Total Hotel	158
B.19	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Kamar Hotel Bintang 1	159
B.20	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Objek Wisata Alam	160
B.21	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Kamar Hotel Bintang 3	161
B.22	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Kamar Hotel Bintang 4	162
B.23	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Total Kamar Hotel	163
B.24	Hasil Hubungan Antara Jumlah Wisatawan Nusantara Dengan Jumlah Restoran, Rumah Makan, dan <i>Cafe</i>	164
B.25	Hasil Hubungan Antara Jumlah Wisatawan Mancanegara Dengan Jumlah Total Angkutan Wisata	165
B.26	Hasil Halaman Awal Dari <i>Website Dashboard</i>	165

B.27 Hasil Halaman Awal Dari <i>Website Dashboard</i> Setelah Menekan <i>Icon Hamburger</i> . .	166
B.28 Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi Yang Telah Dibuat	166
B.29 Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi <i>Trend</i> Dari Data	167
B.30 Hasil Halaman Dari <i>Website Dashboard</i> Yang Memiliki Submenu Pada Menu Di <i>Sidebar</i>	167
B.31 Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi Distribusi Data	168
B.32 Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi <i>Correlation Plot</i>	168
B.33 Hasil Halaman Dari <i>Website Dashboard</i> Untuk Menampilkan Visualisasi <i>Scatterplot</i>	169

DAFTAR TABEL

1.1	Jumlah Wisatawan Mancanegara Tahun 2018 dan 2019 ¹	3
2.1	Contoh Tabel Persentase Tingkat Penghunian Kamar Hotel Tahun 2010-2012	8
2.2	Contoh Data Pengunjung Yang Terdapat Di Open Data Jabar	8
2.3	Contoh Dari Matriks Korelasi	14
3.1	Data Terkait Jumlah Objek Wisata Di Setiap Kota Berdasarkan Jenis Objek Wisata	18
3.2	Hasil Tabel Setelah Dilakukan Proses Nomor 1 Dan 2	19
3.3	Hasil Tabel Setelah Dilakukan Proses Nomor 3	19
3.4	<i>Dataset</i> Sebelum Dipecah Menjadi <i>Dataset-Dataset</i> Kecil	20
3.5	<i>Dataset</i> Sesudah Dipecah Menjadi <i>Dataset-Dataset</i> Kecil	20
3.6	Data Yang Digunakan Untuk Menguji Shapiro-Wilk	21
3.7	<i>Dataset</i> sesudah dilakukan proses pengisian <i>missing value</i> (data yang memiliki '**' merupakan data yang sebelumnya merupakan <i>missing value</i>	23
3.8	<i>Dataset</i> sesudah dilakukan proses pengisian <i>missing value</i> (data yang memiliki '**' merupakan data yang sebelumnya merupakan <i>missing value</i>	24
3.9	<i>Dataset</i> sesudah dilakukan proses pengisian <i>missing value</i> (data yang memiliki '**' merupakan data yang sebelumnya merupakan <i>missing value</i>	25
3.10	<i>Dataset</i> Setelah Diolah Menjadi Bentuk Tabular	26
3.11	Bentuk <i>Dataset</i> Setelah Di <i>Import</i> Ke Dalam R	26
3.12	Bentuk <i>Dataset</i> Setelah Mengubah Nama Kolom	27
3.13	<i>Dataset</i> Hasil Penggabungan Masing-Masing <i>Dataset</i> Persentase Hunian	27
4.1	5 Data Teratas Setelah Dilakukan Penggabungan	66
4.2	Deskripsi Kolom <i>Dataset</i> Gabungan	69
4.3	Deskripsi Kolom <i>Dataset</i> Persentase Hunian	69
4.4	3 Bulan Tertinggi Di Setaip Tahunnya	71
4.5	Daftar Variabel Yang Berpengaruh Terhadap Jumlah Wisatawan Nusantara Dan Mancanegara	72
C.1	30 Data Dari <i>Dataset</i> Persentase Tingkat Hunian Kamar Hotel	171
C.2	30 Baris Dari <i>Dataset</i> Gabungan (1)	172
C.3	30 Baris Dari <i>Dataset</i> Gabungan (2)	173
C.4	30 Baris Dari <i>Dataset</i> Gabungan (3)	174

DAFTAR KODE PROGRAM

A.1	PenyiapanData.R	105
A.2	Eksplorasi.R	109
A.3	app.R	126

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pariwisata adalah berbagai macam kegiatan wisata dan didukung berbagai fasilitas serta layanan yang disediakan oleh masyarakat, pengusaha, pemerintah, dan pemerintah daerah¹. Fasilitas serta layanan yang mendukung kegiatan wisata dapat disebut juga sebagai industri pariwisata. Industri pariwisata adalah kumpulan usaha pariwisata yang saling terkait dalam rangka menghasilkan barang dan/atau jasa bagi pemenuhan kebutuhan wisatawan dalam penyelenggaraan pariwisata². Industri pariwisata meliputi akomodasi untuk pengunjung, kegiatan layanan makanan dan minuman, angkutan penumpang, agen perjalanan wisata dan kegiatan reservasi lainnya, kegiatan budaya, dan kegiatan olahraga dan hiburan.

Pada tahun 2020, terdapat wabah penyakit pernapasan yaitu Coronavirus Disease atau biasa disingkat COVID-19 yang menyebar di seluruh dunia. COVID-19 adalah sebuah penyakit pernapasan menular yang disebabkan oleh virus SARS-CoV-2³. Penyakit ini dapat menular melalui udara atau melalui percikan *saliva* yang dihasilkan oleh batuk atau bersin dan mendarat di mata, hidung, atau mulut. Akibat dari COVID-19 yang telah menjadi penyakit pandemik, Pemerintah Indonesia membuat kebijakan-kebijakan untuk mengatasi penyebaran pandemik COVID-19, antara lain:

- Berdiam diri di rumah (*Stay at Home*);
- Pembatasan sosial (*Social Distancing*);
- Pembatasan fisik (*Physical Distancing*);
- Penggunaan Alat Pelindung Diri (APD) seperti masker;
- Menjaga kebersihan diri (cuci tangan);
- Bekerja dan belajar di rumah (*Work/Study From Home*);
- Menunda semua kegiatan yang mengumpulkan orang banyak;
- Pembatasan Sosial Berskala Besar (PSBB);
- Pemberlakuan kebijakan *New Normal*.

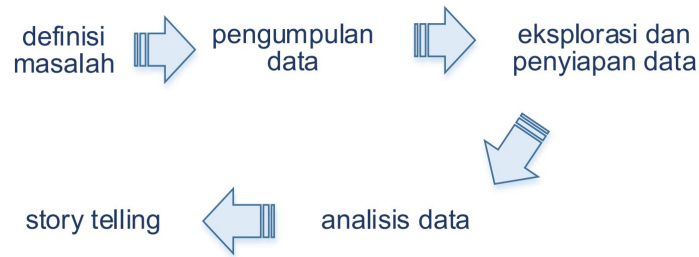
Akibat dari kebijakan-kebijakan yang telah dibuat tersebut membuat masyarakat terpaksa harus beraktivitas dari rumah dan tidak dapat keluar rumah kecuali terdapat aktivitas yang sifatnya darurat. Hal ini menyebabkan masyarakat tidak dapat melakukan perjalanan wisata karena dapat menyebabkan kerumunan orang, sehingga industri pariwisata terkena dampak yang sangat besar. Sepanjang tahun 2020, di Jawa Barat, terdapat lebih dari 2,700 industri pariwisata yang terkena dampak dari kebijakan mengatasi penyebaran pandemik COVID-19 dikarenakan tidak adanya kunjungan wisatawan ke tempat-tempat wisata di Jawa Barat. Akan tetapi di tahun 2022, pandemik COVID-19 sudah mulai mereda dan stabil sehingga terdapat beberapa kebijakan pemerintah yang sudah tidak berlaku, seperti bekerja dan belajar dari rumah, menunda semua kegiatan yang mengumpulkan orang banyak, Pembatasan Sosial Berskala Besar dan berdiam diri di rumah. Hal ini tentu kabar baik bagi industri pariwisata karena masyarakat dapat berpariwisata kembali. Pemerintah Jawa Barat pun mengharapkan industri pariwisata dapat kembali bangkit

¹Undang-Undang Republik Indonesia Nomor 10 Tahun 2009

²Undang-Undang Pariwisata Nomor 10 Tahun 2009

³https://www.who.int/health-topics/coronavirus#tab=tab_1, diakses pada 18 Juni 2022

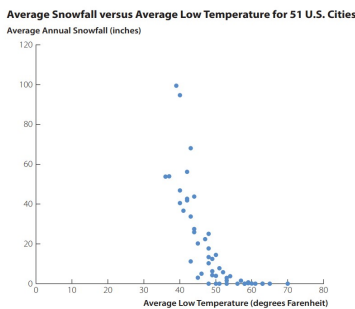
seperti semula. Untuk mencapai hal tersebut diperlukan strategi yang tepat dan untuk menemukan strategi yang tepat dapat memanfaatkan teknik-teknik *data science* untuk mempelajari data-data lampau untuk memperoleh *insight*.



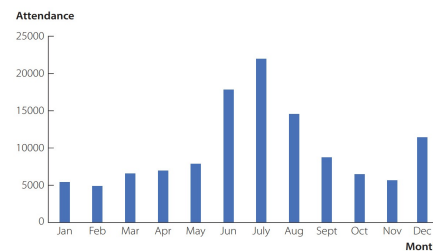
Gambar 1.1: Tahapan *Data Science*[1]

Data science adalah sebuah cabang ilmu yang mengkombinasikan koding, statistika dan bisnis dengan tujuan akhir untuk memperoleh *insight* dari data lampau[1]. Untuk mendapatkan *insight* yang diinginkan, perlu melalui beberapa tahap yang dapat dilihat pada Gambar 1.1. Tahap-tahap tersebut adalah pendefinisian masalah, pengumpulan data, eksplorasi dan penyiapan data, analisis data dan terakhir *storytelling*.

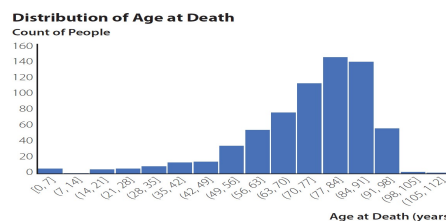
Teknik eksplorasi yang dapat dilakukan untuk mengeksplor data adalah dengan menggunakan statistika deskriptif seperti melihat nilai *mean*, *modus* dan *median*. Teknik lainnya yang dapat dilakukan adalah dengan menggunakan visualisasi seperti *scatter plot*, *bar plot*, *histogram* (lihat Gambar 1.2) dan berbagai macam visualisasi lainnya.



(a) Contoh *Scatter Plot*



(b) Contoh *Bar Plot*



(c) Contoh Histogram

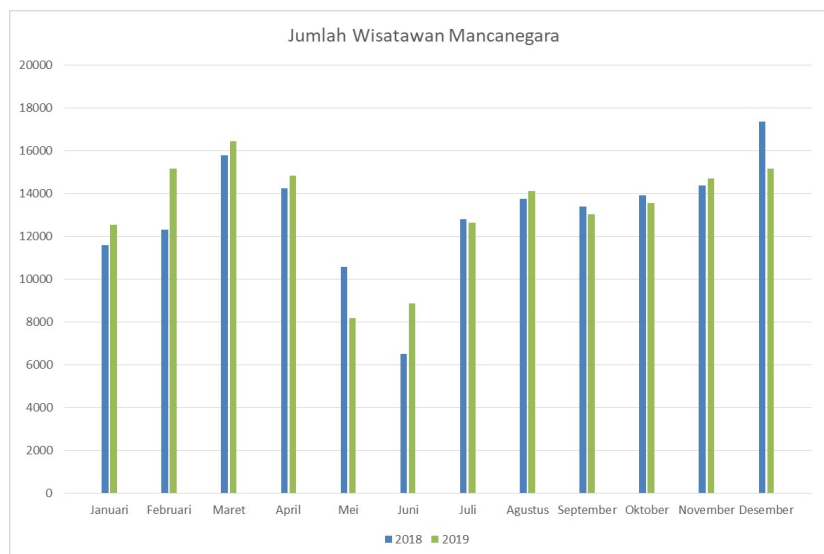
Gambar 1.2: Contoh-Contoh Visualisasi Data[2]

Pada skripsi ini dilakukan analisis dan membuat visualisasi dari data-data terkait industri pariwisata di masa lampau, guna memperoleh *insight* untuk menentukan strategi yang tepat untuk membangkitkan kembali industri pariwisata di Jawa Barat. Sebagai contoh, pada Tabel 1.1 diberikan sebuah data kunjungan wisatawan mancanegara pada tahun 2018 dan 2019. Berdasarkan data tersebut, kita dapat menganalisis data tersebut dengan bantuan visualisasi untuk dapat melihat *insight* dari data tersebut. Dari hasil visualisasi yang ditunjukkan pada Gambar 1.3, diperoleh *insight* bahwa kunjungan wisatawan mancanegara paling banyak terdapat pada bulan desember dan bulan maret. Dari hasil *insight* ini, para pelaku industri wisata dapat menentukan

strategi yang tepat dan bersiap-siap untuk kunjungan wisatawan di bulan maret dan desember seperti di tempat penginapan, mereka dapat menambah karyawan dan ketersediaan tempat tidur untuk memaksimalkan pelayanan penginapan serta memberikan promosi-promosi untuk menarik wisatawan. Analisis dan visualisasi yang dilakukan terhadap *dataset* dibuat dengan menggunakan *R*. Kemudian dibuat juga sebuah perangkat lunak berupa *website dashboard* dengan menggunakan *RShiny* untuk menampilkan visualisasi dari *dataset* yang sudah dibuat.

Tabel 1.1: Jumlah Wisatawan Mancanegara Tahun 2018 dan 2019⁴

Jumlah Wisatawan		
	Tahun	
Bulan	2018	2019
Januari	11,600	12,529
Februari	12,302	15,172
Maret	15,793	16,440
April	14,248	14,830
Mei	10,571	8,168
Juni	6,493	8,881
Juli	12,814	12,645
Agustus	13,766	14,129
September	13,399	13,028
Oktober	13,918	13,569
November	14,364	14,715
Desember	17,375	15,159



Gambar 1.3: *Clustered Barplot* Jumlah Wisatawan Mancanegara

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada skripsi ini adalah:

1. Bagaimana mengumpulkan data terkait industri pariwisata di Jawa Barat?
2. Bagaimana menentukan variabel yang berpengaruh terhadap industri pariwisata?

⁴Data diperoleh dari <https://jabar.bps.go.id/subject/16/pariwisata.html#subjekViewTab3>, pada tanggal 24 September 2022

3. Bagaimana melakukan analisis data yang tepat untuk data industri pariwisata?
4. Bagaimana teknik visualisasi data yang tepat untuk memvisualisasikan hasil analisis data?
5. Bagaimana menentukan strategi bisnis dari *insight* yang sudah diperoleh?
6. Bagaimana membangun perangkat lunak untuk menampilkan hasil visualisasi?

1.3 Tujuan

Tujuan yang ingin dicapai dari skripsi ini adalah:

1. Melakukan studi lapangan dalam bentuk pengumpulan data terkait industri pariwisata di Jawa Barat dari Biro Pusat Statistik Jawa Barat dan *Open Data Jabar*.
2. Mempelajari variabel yang mempengaruhi industri pariwisata
3. Menganalisis data industri pariwisata.
4. Menampilkan visualisasi data yang tepat untuk hasil analisis data.
5. Membuat strategi bisnis berdasarkan *insight* yang telah didapatkan.
6. Membangun perangkat lunak berupa *website dashboard* untuk menampilkan hasil visualisasi menggunakan *Rshiny*.

1.4 Batasan Masalah

Batasan dan asumsi yang digunakan dalam penelitian ini adalah:

1. Data yang digunakan adalah data yang diperoleh dari situs Open Data Jabar, situs Badan Pusat Statistik Jawa Barat, dan Dinas Pariwisata dan Kebudayaan Jawa Barat.
2. Data yang digunakan untuk analisis dan visualisasi berada pada rentang tahun 2014 - 2021.
3. Data persentase tingkat hunian hotel yang digunakan berada pada rentang tahun 2010 - 2021.

1.5 Metodologi

Metodologi pada penelitian ini adalah sebagai berikut:

1. Studi Literatur
Pencarian referensi dan sumber-sumber lainnya yang dapat digunakan sebagai acuan dalam pembahasan mengenai industri pariwisata, analisis, dan visualisasi.
2. Pengumpulan Data
Melakukan pengumpulan data terkait industri pariwisata melalui situs Open Data Jabar, Badan Pusat Statistik Jawa Barat, dan Dinas Pariwisata dan Kebudayaan Provinsi Jawa Barat.
3. Penyiapan Data
Melakukan penyiapan data yang sudah diperoleh agar dapat mudah di analisis dan di visualisasikan.
4. Analisis dan Visualisasi Data
Melakukan analisis dan visualisasi terhadap data industri pariwisata untuk mendapatkan sebuah *insight*.
5. Membangun *Website Dashboard*
Membangun *website dashboard* untuk menampilkan hasil visualisasi yang sudah dilakukan.
6. Penulisan Dokumen
Penulisan dokumen skripsi tentang penelitian yang dilakukan.

1.6 Sistematika Pembahasan

Sistematika penulisan pada skripsi ini adalah sebagai berikut:

1. Bab 1: Pendahuluan, membahas terkait latar belakang masalah yang dibahas, rumusan masalah, tujuan penelitian, batasan dan asumsi, dan metodologi penyelesaian masalah.
2. Bab 2: Landasan Teori, membahas terkait dasar-dasar teori yang digunakan dalam penelitian ini. Pembahasan mencakup industri pariwisata, *dataset* industri pariwisata, *preprocessing data*, uji Shapiro-Wilk, histogram, *mean*, *median*, visualisasi data dan korelasi.
3. Bab 3: *Preprocessing Data* Dan Eksplorasi, membahas terkait penyiapan data sebelum diolah dan eksplorasi yang dilakukan terhadap data.
4. Bab 4: Analisis dan Pencarian *Insight*, membahas terkait penganalisisan terhadap data untuk mendapatkan *insight* yang berguna.
5. Bab 5: Perangkat Lunak, membahas terkait pembangunan *website dashboard* untuk menampilkan hasil visualisasi.
6. Bab 6: Kesimpulan dan saran, membahas kesimpulan yang didapatkan dari hasil analisis yang sudah dilakukan dan saran untuk dapat mengembangkan topik industri pariwisata ini lebih lanjut.