



PROYECTO FINAL

Jorge González Piedra



Introducción

Realizaremos un estudio analizando los datos obtenidos de nuestro aeropuerto en San Francisco en el periodo Jul. 2005- Mar. 2016.

Los datos que estudiaremos son los siguientes.

Dato	Explicación
Activity Period	Periodo de actividad. A cada periodo le corresponde un Mes/Año
Operating Airline	Aerolinea que ha operado el vuelo
Operating Airline IATA Code	Código IATA que ha operado el vuelo
Published Airline	Aerolinea publicada para el pueblo
Published Airline IATA Code	Código IATA publicada para el vuelo
GEO Summary	Tipo de vuelo, doméstico o internacional
GEO Region	Region geográfica del vuelo (US, Europa, Asia....etc)
Activity Type Code	Tipo de actividad
Price Category Code	Tipo de tarifa
Terminal	Terminal del vuelo
Boarding Area	Zona de embarque
Passenger count	Número de pasajeros para el vuelo
Adjusted Activity type code	Tipo de actividad, ajustada
Adjusted Passenger count	Número de pasajeros para el vuelo, ajustado
Year	Año del vuelo
Month	Mes del vuelo



Análisis descriptivo

Realizaremos un análisis descriptivos de todas nuestras variables, mostrando media, desviación estándar y moda para las variables numéricas y la moda para las variables categóricas



Activity period

Periodo de actividad – análisis descriptivo

Total	15007
Media	201045.0733
Desviación estándar	313.3361
Mínimo	200507
Máximo	201603
Moda	200807

El periodo de actividad donde se realizaron más vuelos fue el periodo 200807. Correspondiente a **Julio 2008**



Operating airline / IATA code

Aerolínea operadora– análisis descriptivo

Moda	United Airlines – Pre 07/01/2013 (UA)
-------------	--

Published airline / IATA code

Aerolínea publicada– análisis descriptivo

Moda	United Airlines – Pre 07/01/2013 (UA)
-------------	--

United Airlines es la *aerolínea operadora* y la *aerolínea publicada* con más vuelos realizados



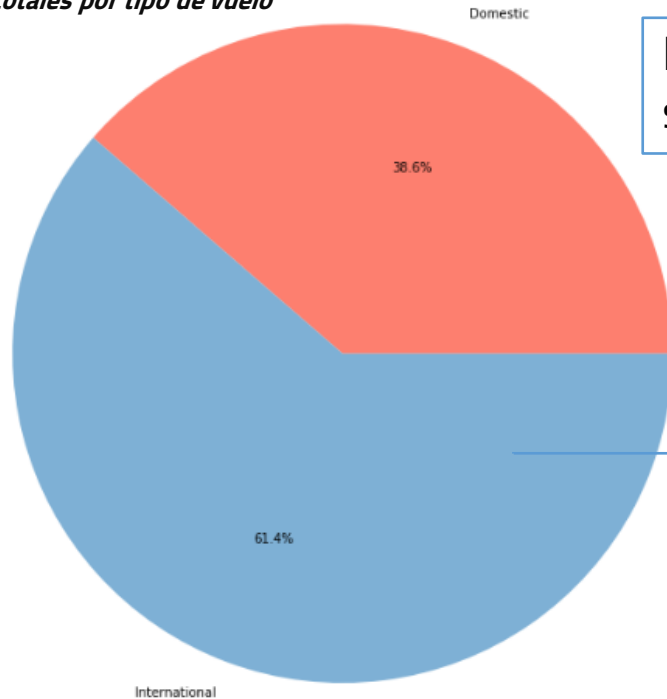
GEO summary

Tipo de vuelo – análisis descriptivo

Moda

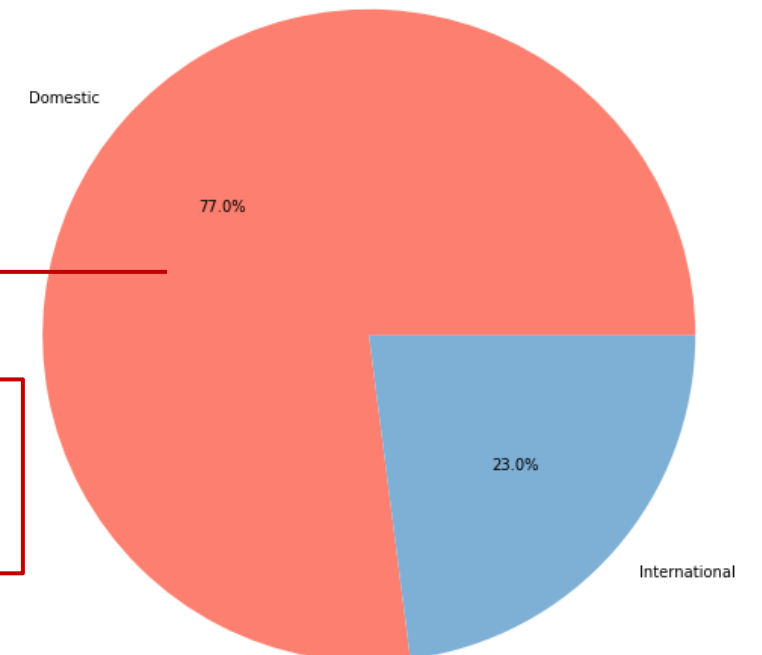
International

*Porcentaje de vuelos
totales por tipo de vuelo*



La mayor parte de los vuelos son **internacionales**.

*Porcentaje de pasajeros
por tipo de vuelo*

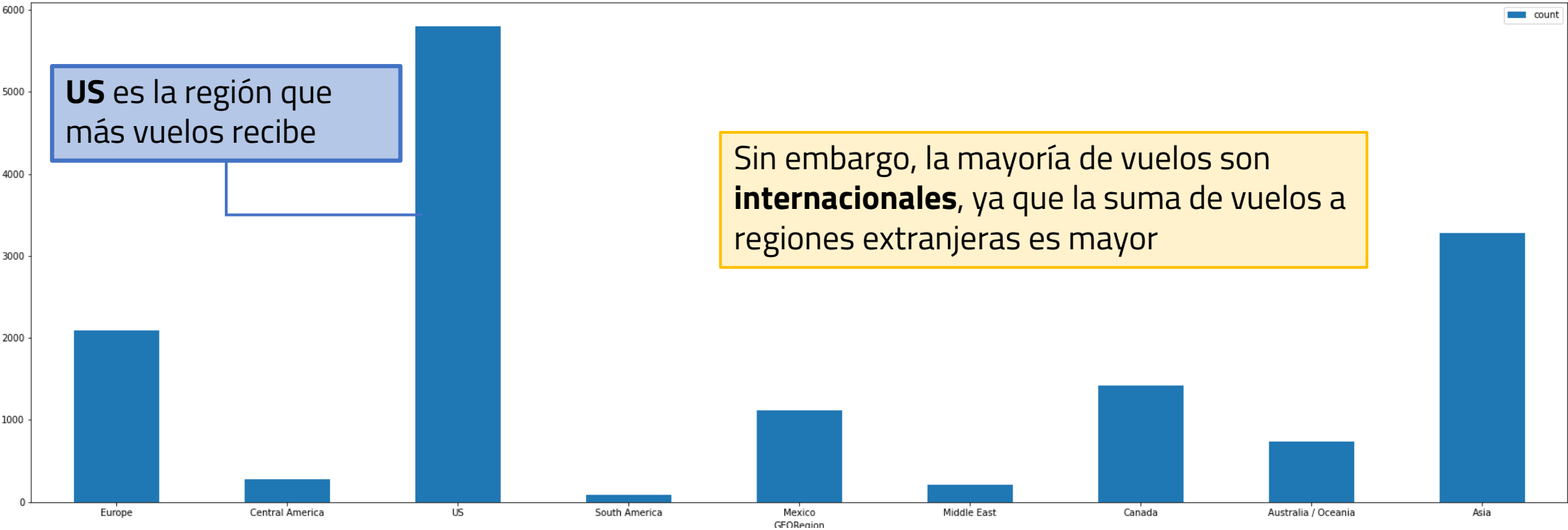


Los vuelos **domésticos** transportan a la mayoría de los pasajeros



GEO region

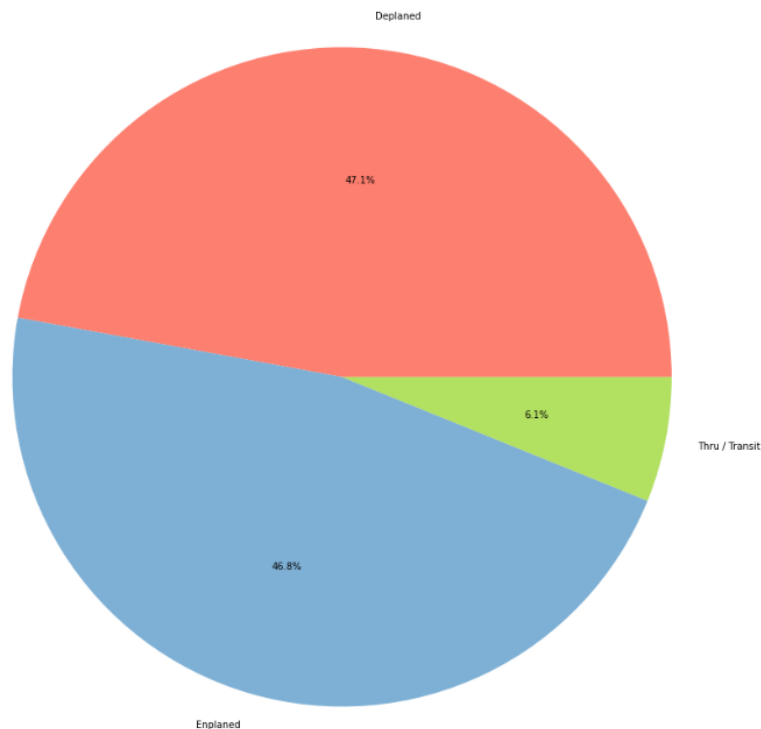
Región– análisis descriptivo	
Moda	US (United States)





Activity type code

Porcentaje de vuelos planificados/desplanificados

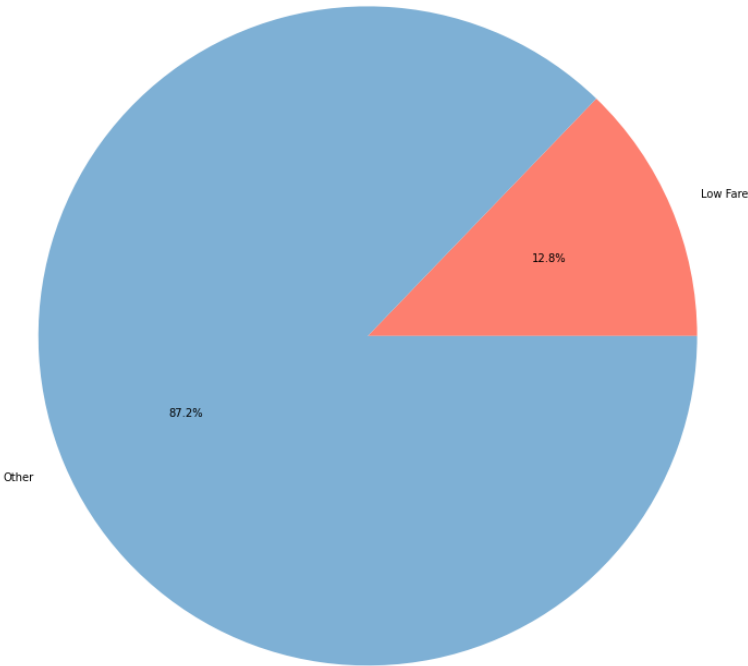


Actividad– análisis descriptivo	
Moda	Deplaned (Desplanificado)



Price category code

Porcentaje de vuelos totales por tipo de tarifa



Tipo de tarifa– análisis descriptivo

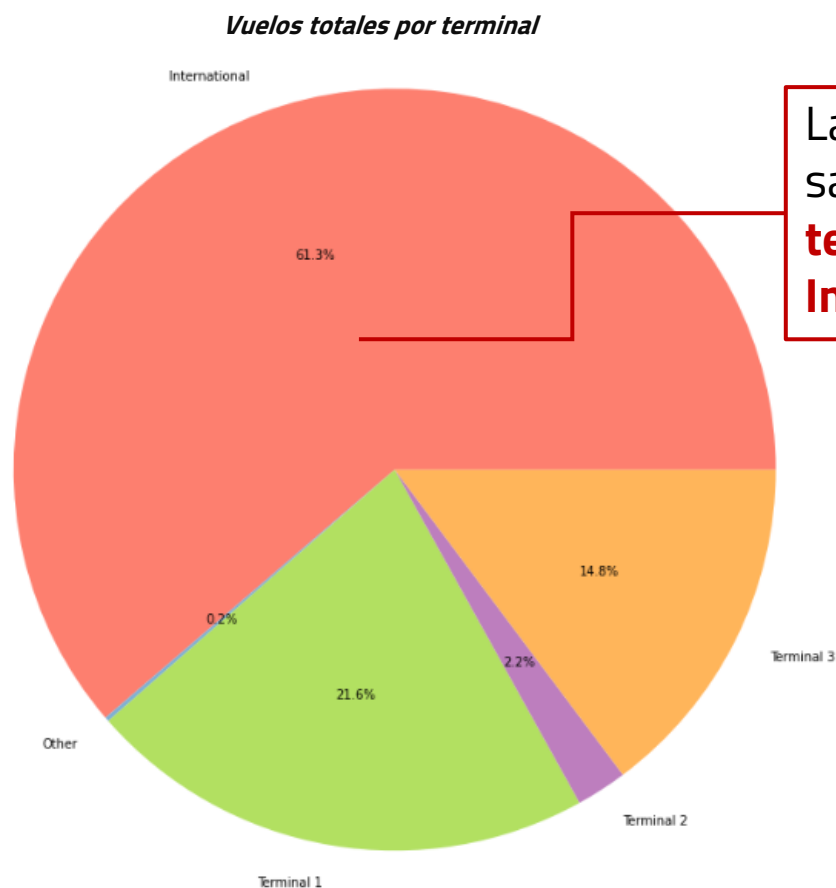
Moda	Other (Otro)
------	--------------



Terminal

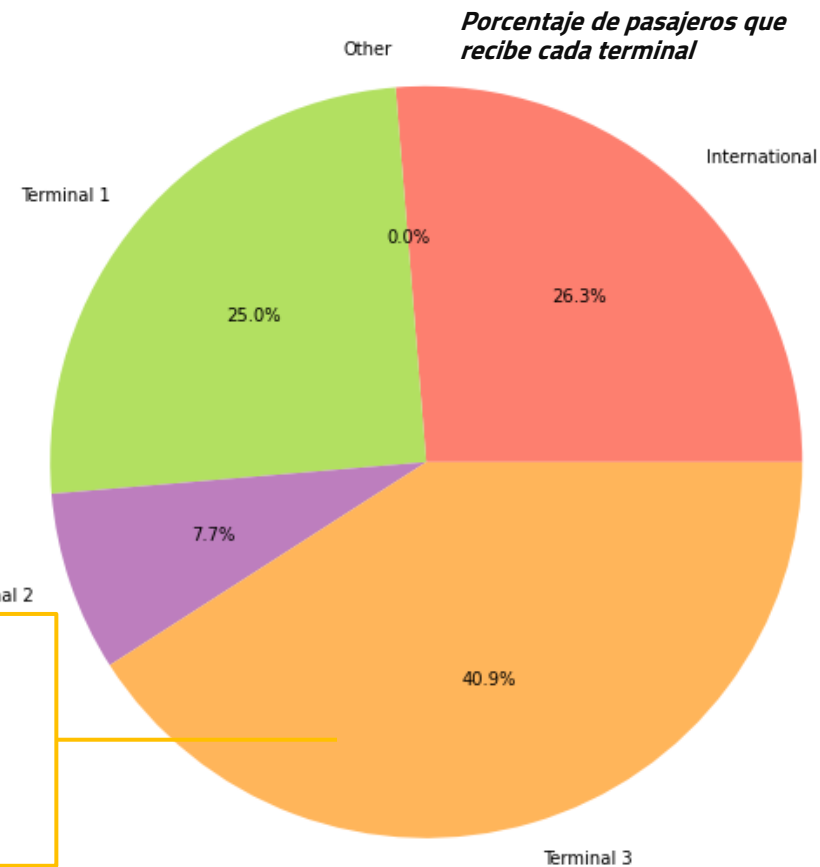
Terminal– análisis descriptivo

Moda	International
------	---------------



La mayoría de vuelos salen desde la **terminal International**

La **terminal 3** es la terminal por la que más pasajeros pasan



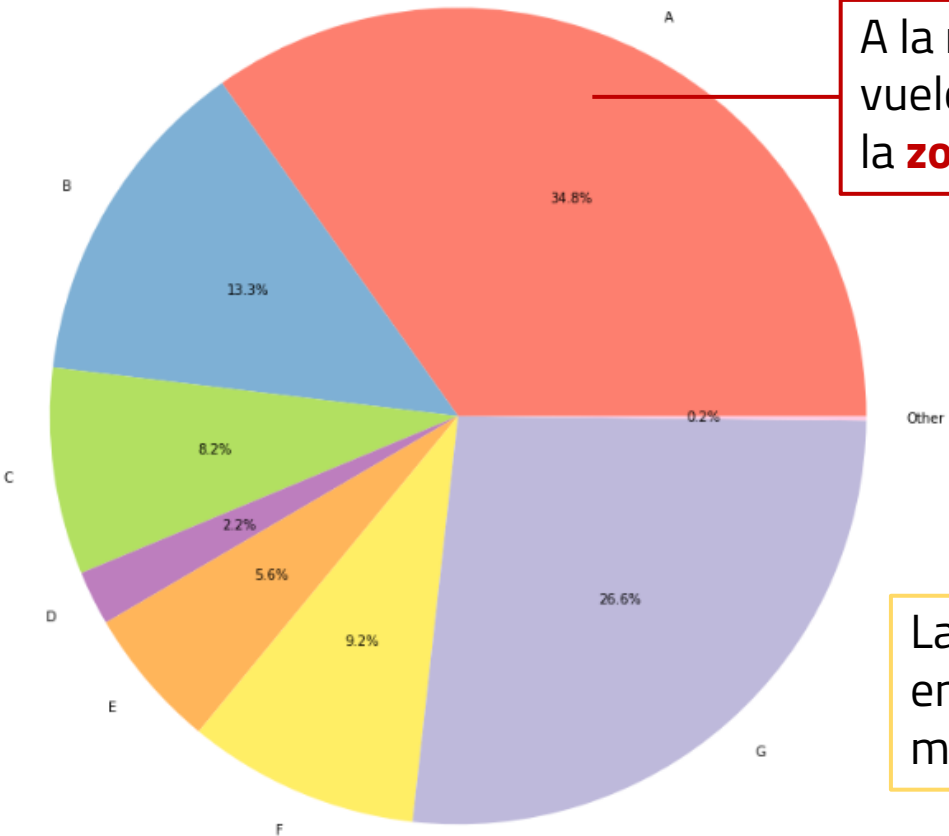


Zona de embarque– análisis descriptivo

Boarding area

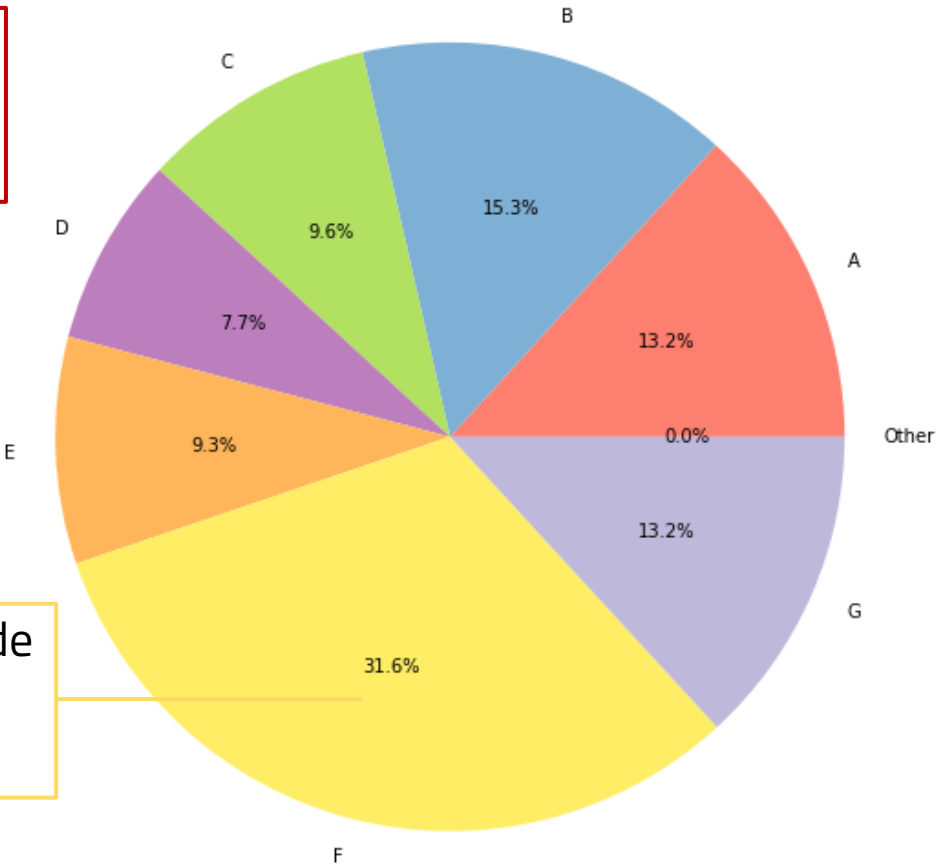
Moda	A
------	---

Vuelos totales por zona de embarque



A la mayor parte de los vuelos se embarca por la **zona A**

Porcentaje de pasajeros que pasan por cada zona de embarque



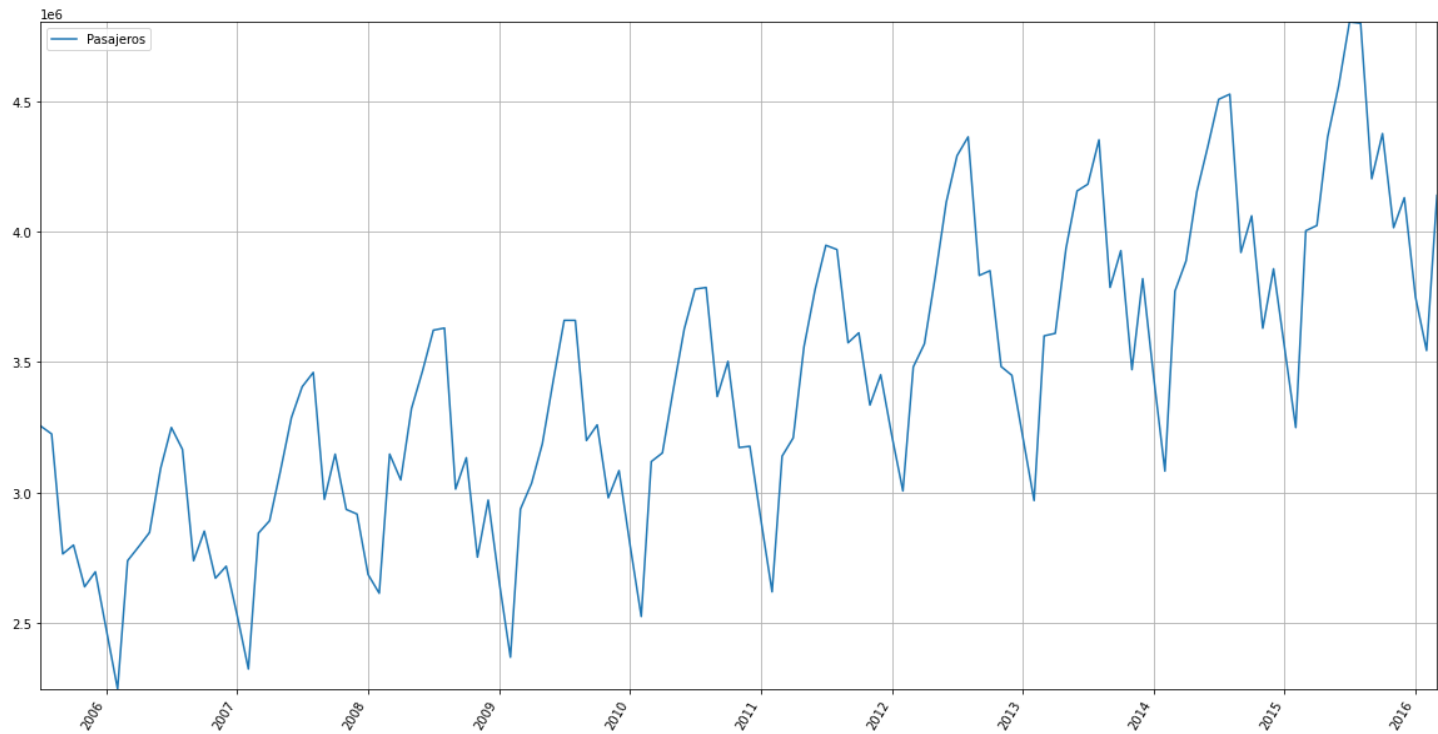
La **zona F** es la zona de embarque por la que mas pasajeros pasan



Passenger count

Periodo de actividad – análisis descriptivo	
Total	15007
Media	29331.9171
Desviación estándar	58284.1822
Mínimo	1
Máximo	659837

Evolución del numero de pasajeros (2005-2016)

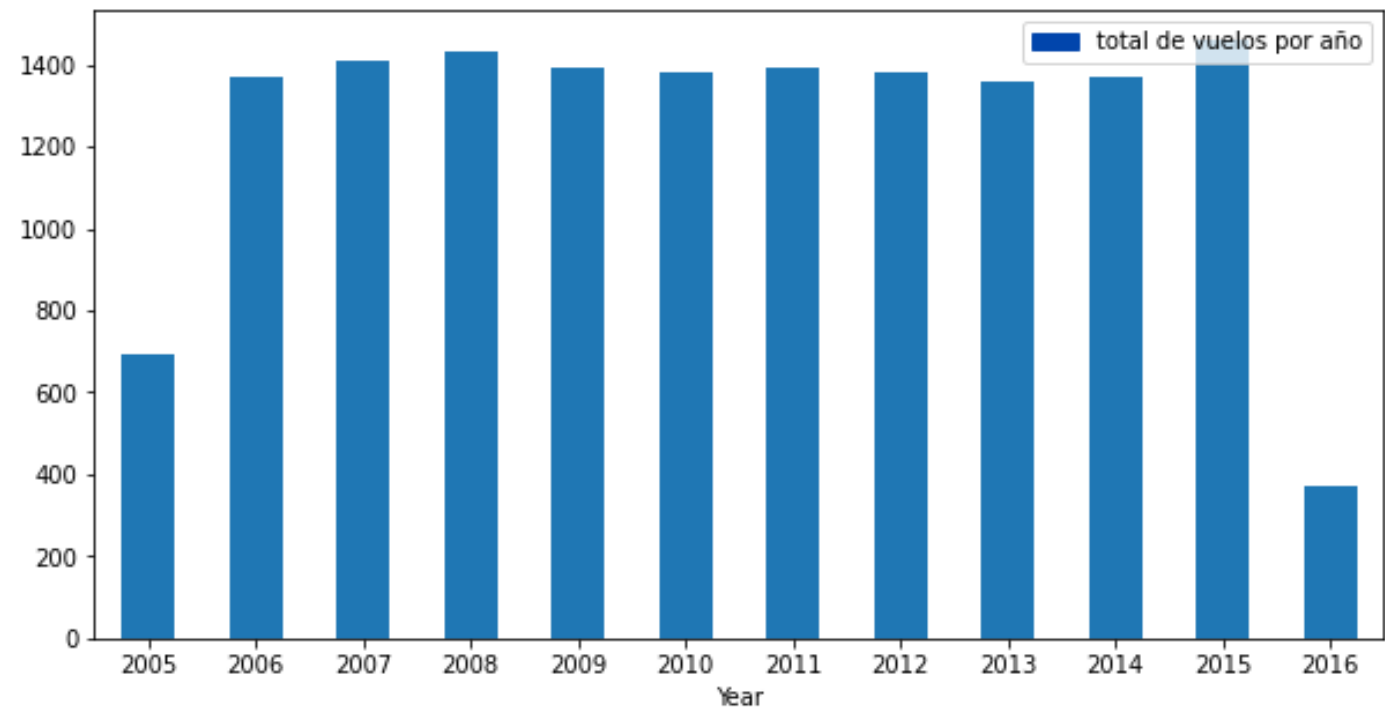




Year

Periodo de actividad – análisis descriptivo	
Total	15007
Media	2010.38522
Desviación estándar	3.13758
Mínimo	2005
Máximo	2016

Numero de vuelos totales por año (2005-2016)

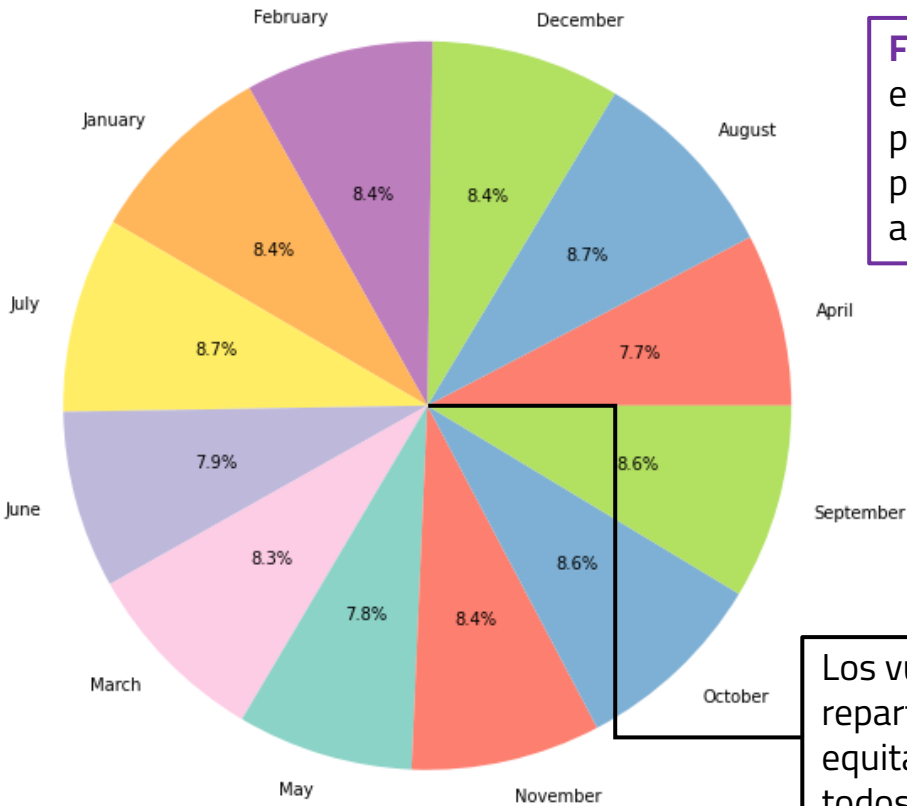




Month

Mes de vuelo– análisis descriptivo	
Moda	Agosto

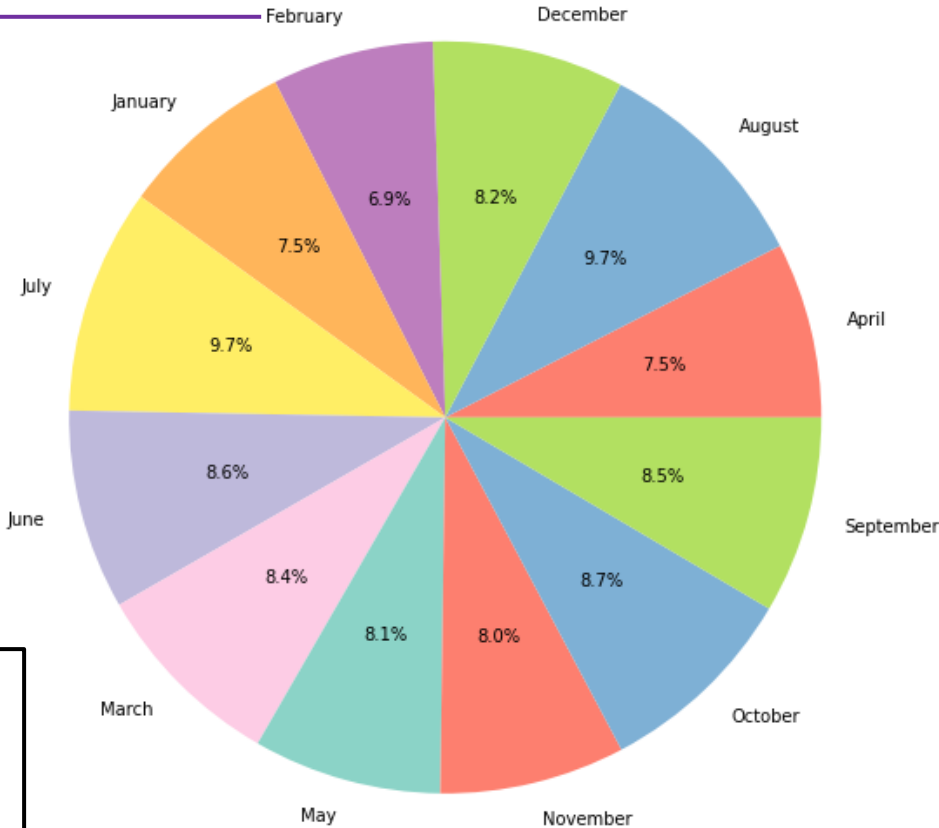
Porcentaje de vuelos realizados cada mes (2005-2016)



Febrero es el mes en el que menos pasajeros pasan por nuestro aeropuerto.

Los vuelos están repartidos muy equitativamente entre todos los meses del año

Porcentaje de pasajeros que han viajado cada mes (2005-2016)





Análisis de correlación

Realizaremos una matriz de correlación y estudiaremos aquellas correlaciones mas fuertes



Análisis de correlación

	ActivityPeriod	OperatingAirline	OperatingAirlineIATACode	PublishedAirline	PublishedAirlineIATACode	GEOSummary	GEORegion	ActivityTypeCode	PriceCategoryCode	Terminal	BoardingArea	PassengerCount	AdjustedActivityTypeCode	AdjustedPassengerCount	Year	Month
ActivityPeriod	1.000000	0.008180	-0.043338	0.009883	-0.010853	0.086100	-0.028145	-0.052450	-0.005754	-0.088284	-0.005123	0.060311	-0.052450	0.059338	0.999940	-0.027142
OperatingAirline	0.008180	1.000000	0.823688	0.968828	0.818090	-0.130956	0.151120	0.100644	-0.096112	0.197959	0.251975	0.185424	0.100644	0.186427	0.008183	-0.000452
OperatingAirlineIATACode	-0.043338	0.823688	1.000000	0.790621	0.919021	-0.131919	0.101535	0.099556	-0.091939	0.208314	0.280294	0.122244	0.099556	0.123170	-0.043257	0.000881
PublishedAirline	0.009883	0.968828	0.790621	1.000000	0.859995	-0.083020	0.108379	0.098414	-0.095284	0.199894	0.275990	0.200862	0.098414	0.201890	0.009641	0.000609
PublishedAirlineIATACode	-0.010853	0.818090	0.919021	0.859995	1.000000	-0.027591	0.008988	0.097818	-0.105385	0.167910	0.312936	0.155368	0.097818	0.156337	-0.010837	0.001689
GEOSummary	0.086100	-0.130956	-0.131919	-0.083020	-0.027591	1.000000	-0.871826	-0.026760	0.411498	-0.574422	0.109553	-0.395743	-0.026760	-0.396856	0.066046	-0.001139
GEORegion	-0.028145	0.151120	0.101535	0.108379	0.008988	-0.871826	1.000000	0.033899	-0.382864	0.509119	-0.121033	0.336113	0.033899	0.336980	-0.028129	0.000949
ActivityTypeCode	-0.052450	0.100644	0.099556	0.098414	0.097818	-0.026760	0.033899	1.000000	0.001004	0.087788	0.087706	-0.071423	1.000000	-0.067804	-0.052364	-0.001523
PriceCategoryCode	-0.005754	-0.096112	-0.091939	-0.095284	-0.105385	0.411498	-0.382864	0.001004	1.000000	-0.102936	0.213485	-0.065047	0.001004	-0.064661	-0.005683	-0.003627
Terminal	-0.088284	0.197959	0.208314	0.199894	0.167910	-0.574422	0.509119	0.087788	-0.102936	1.000000	0.168414	0.429146	0.087788	0.430687	-0.088155	-0.000093
BoardingArea	-0.005123	0.251975	0.280294	0.275990	0.312936	0.109553	-0.121033	0.087706	0.213485	0.168414	1.000000	0.131091	0.087706	0.132147	-0.005109	-0.000581
PassengerCount	0.060311	0.185424	0.122244	0.200862	0.155368	-0.395743	0.336113	-0.071423	-0.065047	0.429146	0.131091	1.000000	-0.071423	0.999941	0.060069	0.000413
AdjustedActivityTypeCode	-0.052450	0.100644	0.099556	0.098414	0.097818	-0.026760	0.033899	1.000000	0.001004	0.087788	0.087706	-0.071423	1.000000	-0.067804	-0.052364	-0.001523
AdjustedPassengerCount	0.059338	0.186427	0.123170	0.201890	0.156337	-0.396856	0.336980	-0.067804	-0.064661	0.430687	0.132147	0.999941	-0.067804	1.000000	0.059096	0.000365
Year	0.999940	0.008183	-0.043257	0.009641	-0.010837	0.066046	-0.028129	-0.052364	-0.005683	-0.088155	-0.005109	0.060069	-0.052364	0.059096	1.000000	-0.030413
Month	-0.027142	-0.000452	0.000881	0.000609	0.001689	-0.001139	0.000949	-0.001523	-0.003627	-0.000093	-0.000581	0.000413	-0.001523	0.000365	-0.030413	1.000000



Correlación más fuertes

Observamos las correlaciones mas fuertes.

Correlación	Coeficiente
GEO Summary (<i>Domestic, International</i>) – Price category (<i>Low fare, Other</i>)	0.4114
GEO Summary (<i>Domestic, International</i>) – GEO Region	-0.8718
GEO Region – Terminal	0.5091
GEO Region – Price category (<i>Low fare, Other</i>)	-0.3828
Passenger count – GEO Region	0.3369
Passenger count - Terminal	0.4306



Variables dicotómicas – Metodo Point-Biserial

Para estudiar de manera más precisa la correlación entre las variables dicotómicas GEO Summary (*Domestic, International*) y Price category code (*Low fare, Other*) con el número de pasajeros, utilizaremos el **método Point-Biserial**.

Correlación	Coefficiente
GEO Summary – Passenger count	-0.0396
Price category – Passenger count	-0.0646



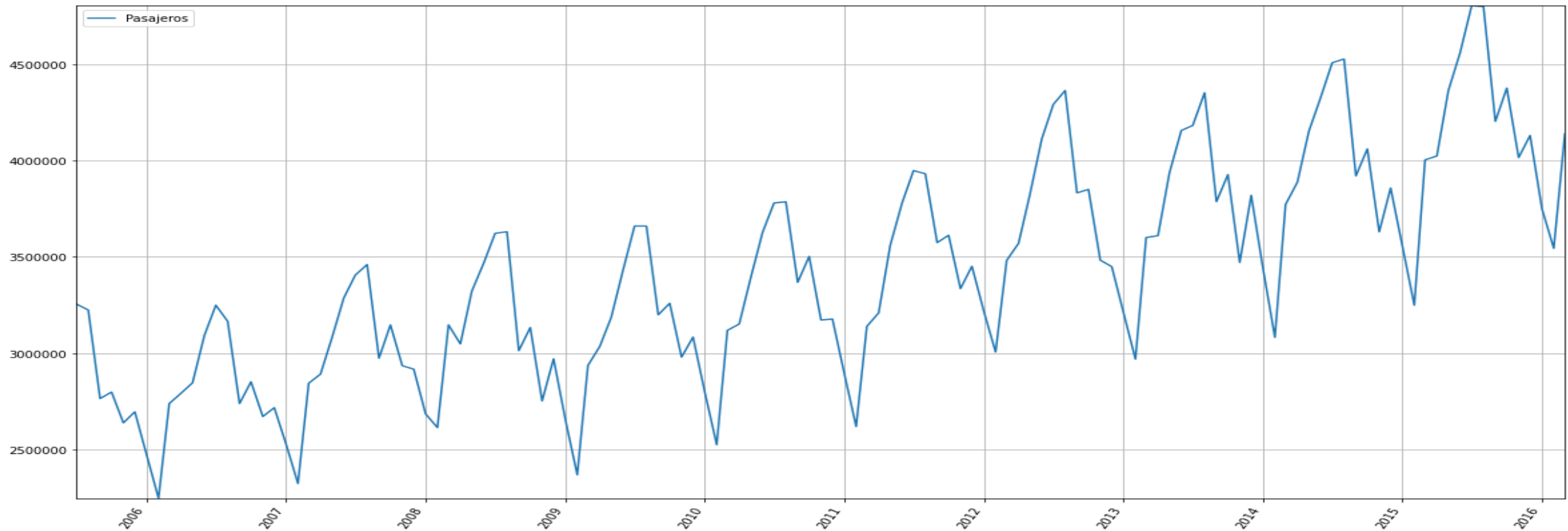
Regresión lineal

Realizaremos una regresión lineal para, a partir de los datos que tenemos, estudiar como variará el número de pasajeros que pasarán por nuestro aeropuerto en el futuro



Regresión lineal

Con un primer análisis visual podemos observar un **patrón** en el número de pasajeros que pasan por nuestro aeropuerto a lo largo del tiempo.





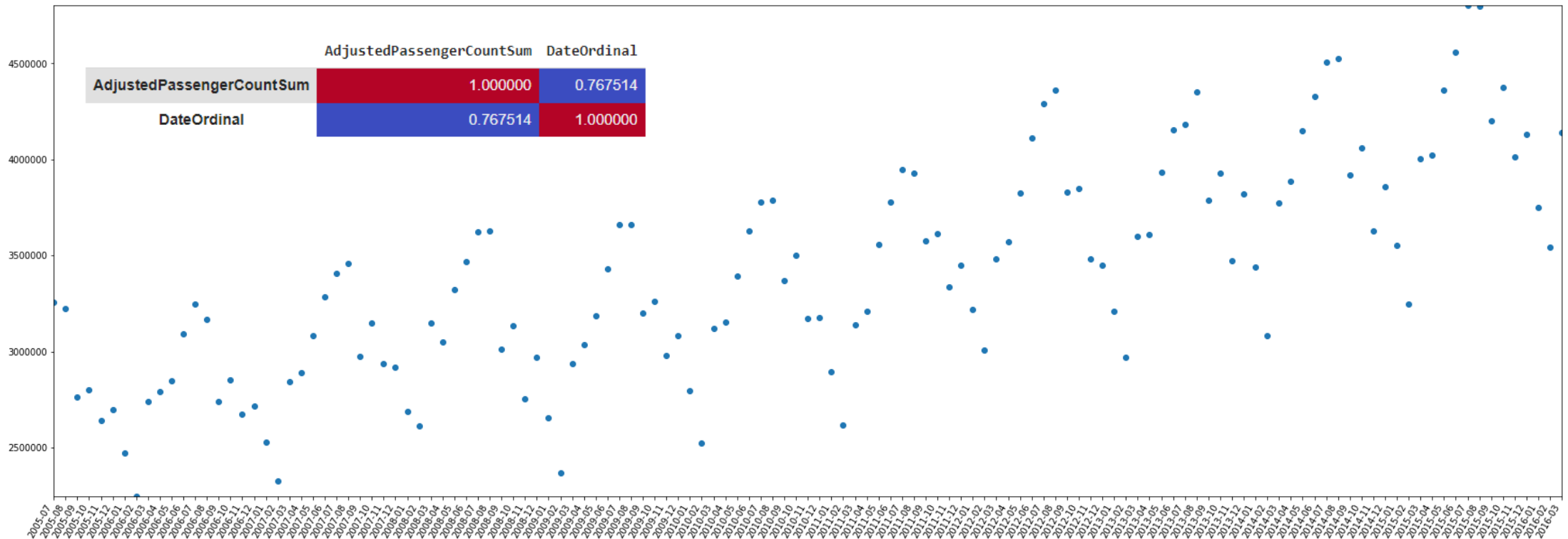
Regresión lineal





Regresión lineal

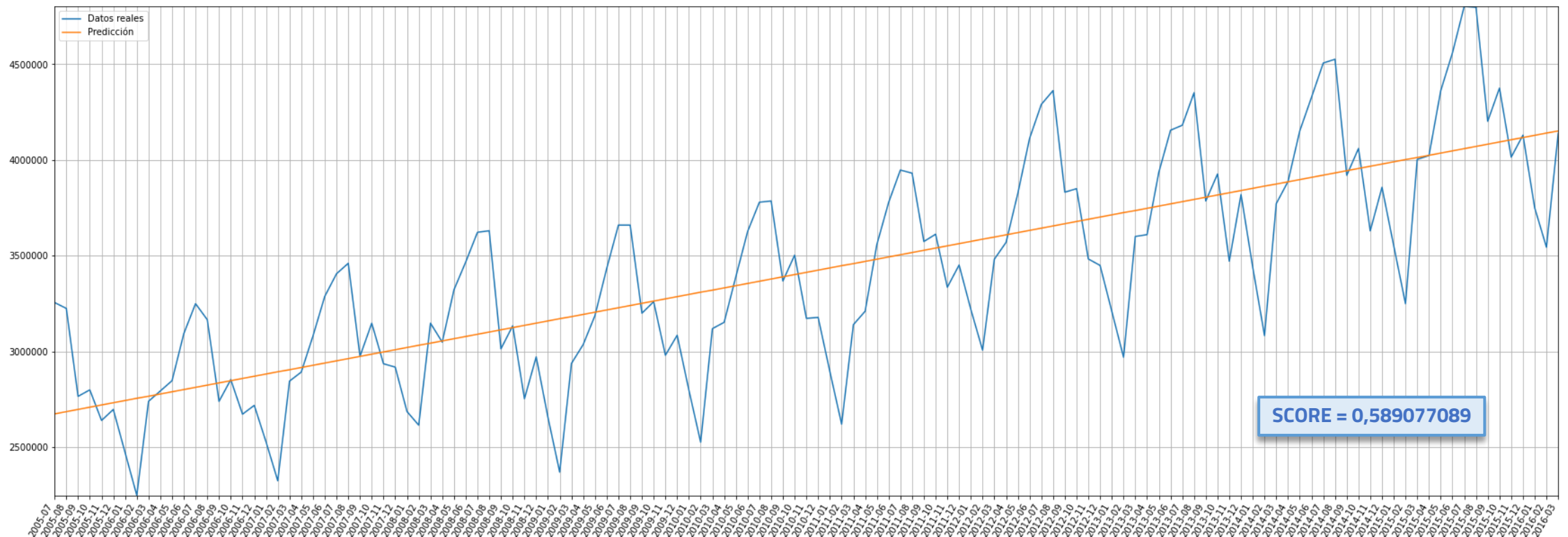
Observamos que existe una **fuerte correlación positiva** entre el número de pasajeros y la fecha, es decir, a medida que aumenta la fecha, crece también el número de pasajeros.





Regresión lineal

Vemos que la tendencia del número de pasajeros ha sido **ascendente** de **forma lineal** desde Julio 2005 hasta Marzo 2016.





Regresión lineal

Si comprobamos la predicción para datos futuros (para todo el año 2016) vemos que esta tendencia ascendente se mantiene.

Sin embargo debemos **tener en cuenta**, mirando los **datos de años anteriores**, que en **Agosto** es previsible que se alcance el **valor máximo** y en los meses posteriores haya un descenso brusco hasta **Febrero** del año siguiente donde se alcanzará el **valor mínimo**.

