

SAD - Clustering

AODA Intelligence

2022-2023

...

Titulación:

Grado en Informática de Gestión y Sistemas de Información

...

3º Curso(2º Cuatrimestre)

Danel Alonso

Oier Arribas

Adrián López

Adrián Cuadrón

15 de mayo de 2023

Índice

1. Introducción	3
2. Algoritmos empleados: Breve Descripción	3
3. Análisis del mejor número de tópicos para nuestra empresa	
US Airways	4
3.1. Valores de la coherencia	5
3.2. Variación de alpha para las K óptimas	6
3.2.1. NEGATIVOS	6
3.2.2. POSITIVOS	11
3.3. Elección del mejor numero de tópicos	14
3.3.1. Negativos	14
3.3.2. Positivos	20
4. Análisis del mejor número de tópicos para la competencia	
UNITED	23
4.1. Valores de la coherencia	23
4.2. Tópicos para las K seleccionadas	24
4.2.1. NEGATIVOS	24
4.2.2. POSITIVOS	26
4.3. Elección del mejor numero de tópicos	27
4.3.1. Negativos	27
4.3.2. Positivos	27
5. Decisiones para crear la historia	28
5.1. Preproceso	28
5.2. Generación de gráficos	28
5.2.1. Gráfico 1	28
5.2.2. Gráfico 2	29
5.2.3. Gráfico 3	30
5.2.4. Gráfico 4	31
5.2.5. Gráfico 5	32
5.2.6. Gráfico 6	33
5.2.7. Gráfico 7	34
5.2.8. Gráfico 8	35
6. Conclusiones	36

1. Introducción

Para esta segunda entrega del proyecto, deberemos hacer un algoritmo de clustering que permita encontrar las razones que hay detrás las opiniones negativas y positivas. Para ello, la empresa US Airways nos ha contratado para saber cuales son las razones de las reseñas negativas y positivas en Twitter. De esta manera, el jefe de la empresa podrá realizar nuevos cambios y mejoras en su aerolínea. Además, tendrá una buena perspectiva de las cosas que están funcionando bien. Dentro del mercado de las aerolíneas nos interesa saber cuales son los puntos fuertes y débiles de nuestro competidor, United, para poder tener una mejor imagen y tener más clientes que ellos.

2. Algoritmos empleados: Breve Descripción

Hemos empleado los siguientes algoritmos con los siguientes hiper-parámetros.

- **Gensim LDA:**

Gensim LDA es una biblioteca de procesamiento de lenguaje natural especializada en la modelización de temas.

- Hiperparámetros: número de tópicos y alpha
- Link: LDA Model - Gensim

- **Scikitlearn LDA:**

Scikit-learn LDA es una técnica de reducción de dimensionalidad general que se puede aplicar a cualquier conjunto de datos.

- Hiperparámetros: número de tópicos
- Link: LDA Model - Scikitlearn

3. Análisis del mejor número de tópicos para nuestra empresa US Airways

La coherencia y la inercia son dos medidas comunes utilizadas para evaluar el rendimiento del modelo LDA y determinar el número óptimo de tópicos.

El algoritmo de Gensim LDA utiliza la medida de la coherencia y se encarga de medir la cohesión semántica de los documentos. Es decir, mide que tan bien las palabras en un documento están relacionadas entre si y que tan distintivo es ese documento en comparación con otros documentos. Cuanto mayor sea el valor de la coherencia, mejor será el modelo.

El algoritmo de Skitlearn LDA utiliza la medida de la inercia que mide la distancia entre los documentos y las palabras. Es decir, mide la dispersión de los documentos dentro de las palabras y la separación entre las palabras. Cuanto menor sea la inercia, mejor será el modelo.

Hemos hecho un barrido del número de tópicos, desde el 1 al 10. Una vez obtenido los valores de la coherencia para Gensim LDA, representamos en una gráfica lineal la coherencia respecto del número de tópicos. Íbamos a sacar también la inercia para Skitlearn LDA pero el método que nos calculaba la inercia debe de tener un bug ($perplexity(X[, sub_sampling])$). En la gráfica obtenida se debería de apreciar un cambio brusco en la evolución de la coherencia, teniendo la línea representada una forma similar a la de un brazo y su codo. El punto en el que se observa ese cambio brusco nos dirá el número óptimo de tópicos a seleccionar para ese data set.

3.1. Valores de la coherencia

■ NEGATIVOS

Numero de tópicos (K)	1	2	3	4	5	6	7	8	9	10
Valores de coherencia	0,2093	0,2664	0,2532	0,2767	0,3385	0,303	0,2798	0,3301	0,3165	0,3087

Cuadro 1: Valores de la coherencia para la clase negativa



■ POSITIVOS

Numero de tópicos (K)	1	2	3	4	5	6	7	8	9	10
Valores de coherencia	0,3489	0,3962	0,4091	0,3949	0,4012	0,356	0,412	0,4251	0,4057	0,4104

Cuadro 2: Valores de la coherencia para la clase positiva



Para la clase negativa vemos que los puntos de codo, es decir, donde se aprecia un cambio brusco en la evolución de la coherencia, es en $K = 2$, $K = 5$ y $K = 8$. Mientras que para la clase positiva sería en $K = 3$ y $K = 8$. Ahora tendríamos que ir jugando con alpha y beta, para ver cual de esas K es la mas óptima para la clase negativa y positiva.

3.2. Variación de alpha para las K óptimas

Para obtener la K (numero de tópicos) óptima para la clase negativa y positiva, hemos hecho pruebas variando el alpha (para $\alpha = 0,1$ y $\alpha = 0,001$) y estos son los resultados/tópicos que nos han dado:

3.2.1. NEGATIVOS

- $K = 2$

- $\alpha = 0,1$

TOPIC 0

flight, **cancelled**, gate, flightled, get, **cancelled_flightled**, hour, plane, still, **delayed**, delay, americanair, charlotte, service, custo-

mer, **late**, waiting, hr, min, sitting.

TOPIC 1

hold, hour, flight, customer, service, **time**, help, get, phone, call, **bag**, **customer_service**, need, reservation, day, number, trying, one, make, amp.

COHERENCIA: 0.2703223800921759

- $\alpha = 0,001$

TOPIC 0

flight, hold, hour, phone, help, **cancelled**, get, call, reservation, customer, time, cancelled_flightled, flightled, day, make, got, min, please, trying, service.

TOPIC 1

flight, hour, service, customer, **cancelled**, get, plane, gate, time, delayed, **delay**, customer_service, waiting, **bag**, late, never, still, hr, today, airline.

COHERENCIA: 0.2796399973612516

- $K = 5$

- $\alpha = 0,1$

TOPIC 0

flight, cancelled, flightled, **cancelled_flightled**, get, hour, day, please, help, **delayed**, **late**, need, would, miss, one, amp, clt, bag, like, plane.

TOPIC 1

hold, flight, time, hour, call, reservation, phone, agent, back, need, hr, thanks, line, still, say, unacceptable, get, flightr, help, **late_flightr**.

TOPIC 2

service, customer, customer_service, gate, flight, hour, still, help, back, waiting, **seat**, plane, get, fee, told, sitting, min, americanair, want, ticket.

TOPIC 3

flight, hour, airline, phone, hold, get, **bag**, service, put, customer, weather, would, **lost**, amp, baggage, cancelled, call, bad, one, minute.

TOPIC 4

flight, cancelled, time, customer, delay, problem, **cancelled_flightled**, flightled, service, delayed, charlotte, **booking_problem**, booking, min, minute, tco, hotel, bag, waiting, take.

COHERENCIA: 0.24231860459070007

- $\alpha = 0,001$

TOPIC 0

flight, hold, hour, help, cancelled, get, phone, minute, trying, need, call, time, reservation, **cancelled_flightled**, flightled, bag, one, please, line, someone.

TOPIC 1

flight, hour, cancelled, **time**, gate, min, thanks, **waiting**, plane, flightled, cancelled_flightled, charlotte, would, sitting, delayed, still, flighted, cancelled_flighted, amp, luggage.

TOPIC 2

flight, service, customer, **customer_service**, get, hour, call, **bag**, worst, ever, cancelled, airline, late, never, first, late_flihtr, flightr, seat, americanair, class.

TOPIC 3

flight, please, clt, **delayed**, get, phone, number, one, plane, delay, weather, issue, people, **miss**, **gate**, tell, staff, tco, update, day.

TOPIC 4

hr, tco, day, dca, waive, phone, customer, flt, gate, mile, **service**, number, americanair, sure, still, dividend_mile, destination, dividend, update, sent.

COHERENCIA: 0.3092330079994777

- $K = 8$

- $\alpha = 0,1$

TOPIC 0

customer, phone, service, customer_service, call, get, help, mile, back, hold, late, flightr, **late_flightr**, min, day, since, **bag**, go, trying, hour.

TOPIC 1

flight, hour, gate, delayed, **waiting**, cancelled, delay, min, hr, plane, crew, late, told, sitting, **seat**, flighted, cancelled_flighted, still, amp, customer.

TOPIC 2

flight, cancelled, hour, problem, hold, cancelled_flightled, flightled, **booking_problem**, booking, get, even, last, americanair, minute, charlotte, still, ticket, bag, put, airport.

TOPIC 3

flight, cancelled, flightled, cancelled_flightled, help, hold, someone, need, please, charlotte, plane, tell, make, dca, hour, **wait**, **long**, clt, going, **treated**.

TOPIC 4

flight, got, today, time, need, hour, like, still, **bag**, airport, plane, tco, charlotte, stuck, back, **website**, another, pay, first, together.

TOPIC 5

flight, get, **cancelled**, use, fee, home, time, number, change, amp, want, two, phone, waive, reservation, **delay**, computer, hr, voucher, **baggage**.

TOPIC 6

service, customer, **customer_service**, hold, make, minute, airline, help, reservation, take, care, email, great, nc, carry, amp, get, complaint, hr, travel.

TOPIC 7

never, time, ever, airline, fly, **bad**, really, worst, **americanair**, **weather**, hour, day, thanks, couldnt, called, trip, joke, phone, mean, sure.

COHERENCIA: 0.33995289033600506

- $\alpha = 0,001$

TOPIC 0

flight, cancelled, flightled, cancelled_flightled, hour, **cancelled flighted**, flighted, charlotte, hr, get, delay, got, bag, today, need, **luggage**, told, **late**, contact, hold.

TOPIC 1

flight, **phone**, **call**, day, late, get, one, tco, someone, americanair, back, really, help, talk, gave, getting, trying, keep, number, person.

TOPIC 2

flight, please, help, problem, get, **delayed**, time, need, **number**, **gate**, issue, phone, plane, booking, **booking_problem**, connection, update, delay, miss, hour.

TOPIC 3

flight, hour, **waiting**, bag, **gate**, work, hotel, home, dca, sitting, still, get, plane, **delayed**, philadelphia, baggage, today, vacation, thank, amp.

TOPIC 4

hour, hold, time, reservation, minute, min, get, back, wait, hr, waiting, want, plane, sitting, make, put, anyone, unacceptable, **money**, airport.

TOPIC 5

weather, know, check, customer, bad, americanair, need, like, call, got, service, would, never, **staff**, since, flying, **bag**, ticket, care, airline.

TOPIC 6

service, customer, **customer_service**, flight, worst, ever, amp, first, airline, time, never, cancelled, still, guy, class, thanks, complaint, date, experience, flightations.

TOPIC 7

min, told, fee, pay, seat, still, waive, **gate**, people, would, use, fly, way, **ice**, **line**, neptune, never, st, desk, delayed.

COHERENCIA: 0.30777798736921036

3.2.2. POSITIVOS

- $K = 3$

- $\alpha = 0,1$

TOPIC 0

thank, flight, thanks, get, time, great, made, **help**, much, **service**, today, appreciate, back, amazing, **customer**, home, amp, please, good, week.

TOPIC 1

thanks, flight, great, **service**, thank, gate, customer, agent, **attendant**, get, team, **helpful**, got, tonight, delay, airport, awesome, charlotte, tco, bag.

TOPIC 2

best, thanks, follow, please, crew, flight, phl, amp, today, clt, tonight, service, **job**, great, airway, **flying**, guy, made, usair, day.

COHERENCIA: 0.39434983222637165

- $\alpha = 0,001$

TOPIC 0

flight, thanks, great, thank, **service**, get, phl, tonight, home, **customer**, clt, awesome, today, phx, **team**, amp, attendant, time, gate, good.

TOPIC 1

thanks, flight, amp, help, made, gate, get, today, thank, agent, problem, back, day, **easy**, phone, **booking**, cancelled, time, rescheduled, getting.

TOPIC 2

thank, flight, thanks, best, great, please, service, follow, amazing, **time**, customer, much, crew, **job**, get, appreciate, made, airport, **helpful**, bag.

COHERENCIA: 0.3593144184867189

■ $K = 8$

● $\alpha = 0,1$

TOPIC 0

thanks, get, **flight**, gate, thank, getting, agent, home, **staff**, help, today, americanair, issue, charlotte, **earlier**, people, sure, tonight, airway, thing.

TOPIC 1

thank, flight, amp, great, today, appreciate, clt, airway, **service**, **check**, employee, phl, better, **help**, awesome, like, got, tco, cancelled, yes.

TOPIC 2

great, job, flight, thank, made, crew, super, **plane**, service, get, best, awesome, charlotte, team, **attendant**, delayed, time, **professional**, see, eliz.

TOPIC 3

flight, thanks, **service**, thank, phl, customer, great, **crew**, wonderful, kudos, still, agent, everything, happy, reply, home, people, attendant, get, day.

TOPIC 4

follow, please, best, service, thanks, amazing, customer, **time**, phl, back, appreciate, tonight, great, week, today, **flying**, airport, tco, phx, desk.

TOPIC 5

flight, time, thanks, good, great, gate, made, day, problem, **help**, well, airline, work, flying, team, mind, agent, helpful, easy, despite.

TOPIC 6

thanks, great, made, tco, customer, **service**, guy, flight, please, good, worry, worked, team, fine, helpful, nice, airport, another, attendant, name.

TOPIC 7

thank, flight, much, ok, air, time, delay, great, today, problem, crew, tomorrow, **bag**, fa, phx, booking, texas, ground, **helpful**, response.

COHERENCIA: 0.45731929234848206

- $\alpha = 0,001$

TOPIC 0

thanks, flight, time, great, work, thank, gate, today, job, customer, service, bag, phx, get, like, agent, look, clt, please, never.

TOPIC 1

service, flight, thanks, customer, get, great, amp, made, back, agent, gate, desk, waiting, appreciate, helpful, right, well, rebooked, make, miss.

TOPIC 2

flight, thank, phl, thanks, great, attendant, team, get, tonight, home, fa, back, helpful, amazing, today, fantastic, job, helped, airline, totally.

TOPIC 3

tco, thanks, made, problem, thank, clt, best, airport, thx, get, people, customer, crew, great, connection, service, kudos, booking, cancelled, helpful.

TOPIC 4

flight, thank, yes, today, awesome, got, great, americanair, appreciate, guy, home, thanks, job, team, much, good, dca, best, dream, flying.

TOPIC 5

thank, follow, please, best, thanks, great, get, flight, delay, amazing, much, agent, getting, back, airway, earlier, able, gate, check, eliz.

TOPIC 6

thanks, thank, flight, service, amp, time, everything, phl, customer, away, today, day, staff, professional, hoping, catch, good,

week, another, departure.

TOPIC 7

thank, great, help, thanks, crew, made, airline, amp, please, thing, however, texas, keep, work, delayed, flight, air, see, guy, flying.

COHERENCIA: 0.4263197747651563

3.3. Elección del mejor numero de tópicos

3.3.1. Negativos

El número de tópicos que hemos elegido son 8, ya que con 2 o 5 no están presentes las suficientes razones de nuestra compañía, es decir, hemos considerado que con los tópicos 2 y 5 cada tópico hablaba de mas de una cosa en concreto lo cual generaría conclusiones ambiguas. Además, con 8 tópicos es donde conseguimos el mayor número de coherencia, lo cual es una razón importante para seleccionarlo. También hemos probado con diferentes alphas, y hemos comprobado que al cambiar el alpha, la distribución de los tópicos y la coherencia no cambia mucho. Hemos probado con dos alpha un poco separados (0.1 y 0.001) para ver si notamos diferencia, pero no ha sido el caso.

- $K = 8$
 - 0.1 \rightarrow late_flight, wait y delay, booking problem, cancelled flight, baggage lost
 - 0.001 \rightarrow cancelled flight, call person, booking problem, delay , waiting, bag lost, customer service, long lines

Una vez tenemos el numero óptimo de tópicos, lo que hemos hecho ha sido llevarlo a la plantilla del LDA de Scikit-Learn para que nos saque los documentos mas importantes de cada tópico. Gracias a esto, vamos a tener mayor contexto de las razones negativas y cual es la causa del problema. Hemos conseguido lo siguiente:

VUELO CANCELADO/CANCELLED FLIGHT

En esta razón negativa podemos ver que hay una mala organización de la empresa US Airways, ya que hay muchos vuelos que se cancelan, no mantienen al cliente informado y eso hace que los clientes tengan que sufrir muchos

tiempos de espera.

@USAirways thank you so much for asking us to pay \$700 to switch to a different flight so I could spend more time with my dying nephew. // MALA ORGANIZACION

@USAirways 2133. Flight now Cancelled Flightled. Have been rebooked on 2pm flight. If on time take off, I'll have spent 5 extra hours at DCA.

@USAirways #2066. Was on plane from PBI to CLT and knew about the frozen water. Also saw a plane to NYC take off at the gate next door!

@USAirways after 5 flight Cancelled Flightlations we finally get a flight, get to the gate on time & you've given our seats away? Seriously?? //5 VUELOS CANCELADOS Y SIN ASIENTOS DISPONIBLES

@USAirways why did you Cancelled Flight flight 1773 to phl, while delta flew a flight at the same time, on time? you Cancelled Flighted for no reason // VUELO CANCELADO

@USAirways I'm unable to check in for flight 2119 BOS-DCA and haven't heard if flight was Cancelled Flighted/rescheduled. Any updates?? // VUELO CANCELADO Y NO SE MANTIENE INFORMAD AL CLIENTE

PROBLEMAS DE RESERVA/FLIGHT BOOKING PROBLEMS

En esta se razón se puede ver una deficiencia importante a la hora en cuanto al sistema de reservas de vuelo. Podemos ver analizando un poco los Tweets que esto se debe principalmente a la pagina web, concretamente en la eficiencia de esta(muy mala velocidad de conexión y muchas perdidas de la misma, habría que mejorar los servidores que la albergan). Por otra parte parece ser que la gestión de vuelos es muy tosca, cuesta mucho reasignar y cancelar billete. Finalmente en otros Tweets hemos visto que el servicio telefónico(largos tiempos de espera y pocas personas para atender a la gente, esto dificulta mucho también las reservas vía telefónica.

@USAirways I tried reaching out to you guys I'm running out time and patience I just want my flight to be honor and I want a confirmation! // CUESTA MUCHO RESERVAR VUELO

@USAirways ok my 4th try today....giving up at 12.45am after 138 minute hold time. #terrible <http://t.co/Ps9Q9RPSiL> // PAGINA WEB DEFICIENTE, NO SE PUEDE RESERVAR UN VUELO

@USAirways I cant make it to airport, roads are to bad in DFW and no one is answering the phones and the website is down?How do I reschedule // WEB NO FUNCIONA

@USAirways trying to change flight because of airline travel advisory. on hold for 45 mins. Help! // MALA GESTIÓN DE VUELOS

@Usairways thanks for making miss connection, getting rebooked, missing that too, and a miserable day. Started my day @ 4:30am for nothing! // PERDIDA DE CONEXIÓN

RETRASOS/LATE AND DELAYED FLIGHT

En esta razón negativa podemos ver que los clientes se quejan de retrasos en los vuelos, malas condiciones de viaje y tardanza en llegar a los destinos, debido a fallos de la empresa.

@USAirways 30 min delay for mechanics to replace rivets. 30 mins Late Flightr, turns out mechanics had wrong size rivets. 35 more mins. Pathetic

@USAirways US 728 then lost ground power three different times. Not only grounded endlessly; now without air, lights, entertainment. // MALAS CONDICIONES DE VIAJE, TARDANZA EN LLEGAR AL DESTINO

ESPERAS y PUERTAS EMBARQUE/LOGLINES AND GATES

Cambian muchas veces la puerta de embarque y no esta clara cual es. Los agentes de las puertas de embarque. Con los retrasos hay muchas colas de espera.

@USAirways 2 bad planes & 3 gate changes= lots of exercise. #travel-fail let's hope this plane works. I'm tired.

@USAirways oh yay! Now gate agents decide not to deal with upgrade list and just bump up whoever is closest.

PERTENENCIAS PERDIDAS/LOST LUGGAGE

Aquí podemos ver una deficiencia muy importante a la hora de entregar las maletas, con hasta una hora de retraso, si no es que directamente se pierde, esto combinado también con la deficiencia de la pagina web comentada anteriormente que dificulta el seguimiento de las maletas crea un gran descontento entre los clientes. Por no hablar de que a una persona le han robado una cámara que tenia dentro de su maleta.

@USAirways Flight gets in 30 minutes early and all goodwill gone when bags aren't here 45 minutes Late Flight
RETRASO EN LAS MALETAS

@USAirways Baggage Team? Recording tells me my bag still hasn't been found. And delayed baggage link on web conveniently not working. #worst
// WEB NO FUNCIONA y BAG NO ENCONTRADO

@USAirways #PHL why does it take an hour to offload bags, especially after taking 30+ min to get a jetway to get off plane after landing?
RETRASO

@USAirways flt 419. 2+ hrs Late Flight, baggage + 1 more hr. Now I see they delivered my suitcase wet inside & out. #NotHappy // RETRASO EN EL VUELO Y EQUIPAJE MOJADO

@USAirways @AuroraBIZ check your bag, @PHLAirport bag handlers broke

into my suitcase and stole my camera, bag was returned in a trash bag

SERVICIO AL CLIENTE/CUSTOMER SERVICE

Al parecer la gente se queja de que al llamar por teléfono tardan mucho tiempo, incluso horas en conseguir hablar con una persona real, incluso hay veces que ni lo consiguen. También dicen mucho que se desconecta la línea cuando intentan llamar. La comida del avión es muy cara para la calidad que tienen. Las tarifas de las maletas muy caras(\$150) también en comparación con otras aerolíneas(\$50). Al cancelar un vuelo no devuelven el dinero.

Otro de los puntos negativos es que los empleados son groseros y poco profesionales al hacer errores en los vuelos y creando retrasos.

Una de las cosas que hemos averiguado es que hay varios tweets comentando que los domingos hay pocos empleados. Esto también está relacionado con que justo el 22 de febrero, que fue domingo, es cuando más tweets negativos encontramos sobre el nefasto servicio al cliente.

@USAirways after 2 hrs of trying to reach real human, finally did and been on hold for 1:36 hrs for out of synch reservation.

@USAirways #NoService good luck trying to reach a human! Twice sat on hold for over 50 mins... #neveragain

@USAirways The phone line disconnects. How can I be in the queue to be answered when the phone line simply disconnects when I call?

@USAirways tried twice today on hold for 30 min each time. i have things to do so can't live on hold dealing w/ your customer serv failures.

@USAirways FUK U US AIRWAYS WITH YO SHITTY CHICKEN SALAD SANDWICH THAT SO OVERPRICED AND U FUKING MAKE ME WAIT IN A 6 HR LAYOVER FUK U AND // COMIDA CARA E INCOMESTIBLE

@USAirways take a look at 1715. All F seats filled, so someone upgraded their friend instead of following rules. // TODOS LOS ASIENTOS OCUPADOS, MALA SEGURIDAD?

@USAirways then drop your fee! \$150 (\$300RT!!!!) makes no sense; most EU airlines charge \$0. @AirCanada charges a reasonable \$50 // TARIFAS MUY CARAS

@USAirways is not the new @AmericanAir is more like the new @SpiritAirlines. Love AA. Not impressed with the subpar planes and gate agents. // AVIONES Y AGENTES DE EMBARQUE DEFICIENTES

@USAirways Its pretty ridiculous that at PHX sky harbor you have 4 (!) employees working check in on a SUNDAY afternoon. 30 min & counting. // POCOS EMPLEADOS LOS DOMINGOS

@USAirways on the phone over an hour waiting for customer service got wrong flight information cost us 6 hours from trip. Frustrating!!! // SERVICIO AL CLIENTE SE EQUIVOCO EN LA INFORMACIÓN Y 6 HORAS DE RETRASO

@USAirways hey guys! Thanks for answering me and thanks for keeping me on hold for an hour on customer service at 12 pm PST! // ATENCIÓN AL CLIENTE TERRIBLE,RETRASOS

@USAirways your customer service at CLT is terrible!!! My flight was delayed from JFK for 2 hours! Get to CLT to horrid customer service! // ATENCIÓN AL CLIENTE TERRIBLE, RETRASOS

@USAirways I don't like being lied to. No bad weather in dc md va. I've been stuck in CA all day and should've been home by now. You owe me. // LE MIENTEN SOBRE EL TIEMPO (Temporal) Y PODRÍA HABER COGIDO EL VUELO

@US Airways I bought ticket same flights twice and you dont refund money it is big problem other company refund money why dont u pay back // NO SE DEVUELVE EL DINERO AL CANCELAR

@US Airways standby so far for two of us. With snow storm hitting Iowa tomorrow we could be stuck here for two days. // EMPLEADOS GROSEROS

3.3.2. Positivos

Por otra parte, respecto a los mensajes positivos, al contrario de lo que pasaba con los negativos, hemos encontrado de que no hay tantas razones primarias. Por lo tanto, a pesar de que con 8 tópicos el nivel de coherencia era mayor, hemos decidido usar 3 tópicos ya que con 8 hemos visto que muchos de los tópicos que que supuestamente son diferentes hablan de la misma razón. Respecto a los valores de alfa hemos encontrado que con $\alpha=0.001$ los tópicos se distinguen mas entre ellos.

- $K = 3$
 - 0.1 \rightarrow customer, help, service, job, flying, helpful
 - 0.001 \rightarrow service, customer, team / easy, booking / time, job, helpful

Para los positivos también vamos a sacar los documentos mas importantes para ver las causas de las razones. Para ello lo hemos ejecutado en la plantilla del LDA de Scikit-Learn. Hemos obtenido lo siguiente:

TOPIC 0

La gente comenta que los pilotos y azafatas hacen buen trabajo, incluyendo algún nombre que se podrían tener en cuenta para el jefe.

@US Airways shout out to the pilots and FC attendant(Eliz) of US 673. Super strong crosswinds during landing. Eliz did a super job throughout

@US Airways Thks US #1786 2/19 PHL to FLL. Overall 1st class has im-

proved with more food now on PHL to FLL and most crews are super.

@USAirways Painless and effortless flight from Indy to PHL... Our flight attendant, Tory, was fantastic. Give that lady a raise ASAP!

@USAirways First class service on US 769 PHL-MCO today from the flight attendant in F. I didn't catch her name but she was top notch!

TOPIC 1

Están contentos con la gente que trabaja en los aeropuertos por ayudarles así como agente de puertas de embarque y los empleados de reservas.

@USAirways thanks for getting me rescheduled on a direct flight to NOLA tonight in 10 minutes--and hopefully an hour earlier getting there!

@USAirways Shout out to the red-headed gate agent for flt 3389 from DCA to CHS at 7:10 tonight. Didn't get her name but she was great!

@USAirways thanks to the gate agent in State College, PA that was able to get me on an earlier flight AND figure out an earlier connection!

@USAirways Just talked to reservation. Must congratulation to them. Very friendly. Good for usair. The ONLY airline we fly.

TOPIC 2

La gente hace una buena reseña de la gestión que se hace cuando se cancela un vuelo o se retrasa, y reciben buen trato y ayudas.

@USAirways Eyyyy! Cancelled Flightlations, Flight Booking Problemss, re-Flight Booking Problemss, but y'all got me on the same flight out tonight (not tomorrow) & the FC upgrade. Thx!

@USAirways @AmericanAir shout out to Diane at EYW for helping get us home today instead of tomorrow (even if a little Late Flight!)

@USAirways Your CLT baggage crew deserves a #kudos. I had to run to make my connection. And my bags still made it! #CustomerService

@USAirways after missing my flight and reFlight Booking Problems 2x, I just walked onto another flight and my phone was still on the seat!!

@USAirways your ticket agents at gate 4 in Providence airport rocked tonight, especially Kristy, sorry if that is not the correct spelling.

@USAirways with the weather mess in the South, I missed my connection in CLT. Impressed though with the Cust Serv phone reps today.

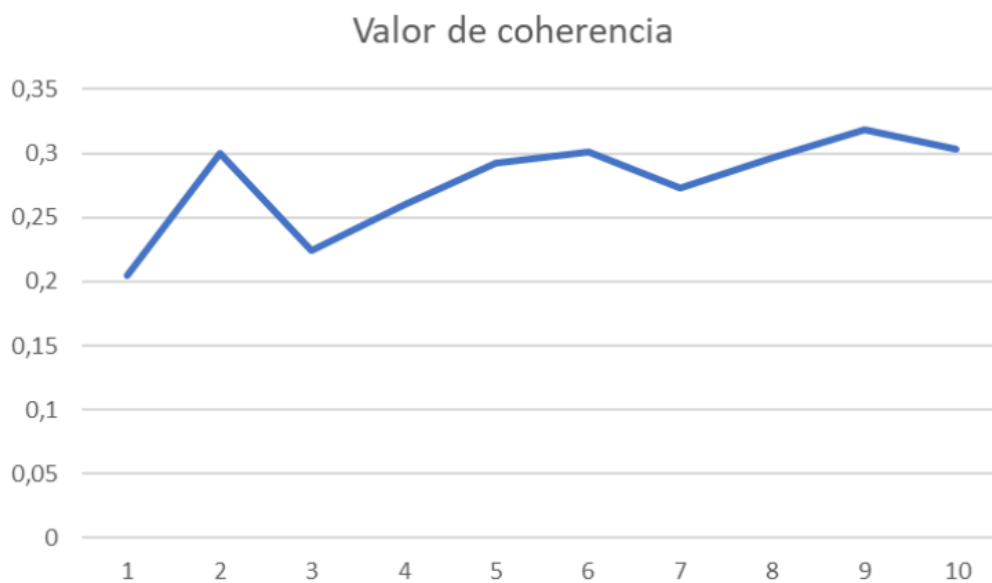
4. Análisis del mejor número de tópicos para la competencia UNITED

4.1. Valores de la coherencia

■ NEGATIVOS

Numero de tópicos (K)	1	2	3	4	5	6	7	8	9	10
Valores de coherencia	0,2046	0,3003	0,2246	0,2605	0,2922	0,301	0,2731	0,2971	0,319	0,303

Cuadro 3: Valores de la coherencia para la clase negativa



■ POSITIVOS

Numero de tópicos (K)	1	2	3	4	5	6	7	8	9	10
Valores de coherencia	0,2613	0,343	0,3721	0,3748	0,389	0,4021	0,4099	0,4091	0,4081	0,4266

Cuadro 4: Valores de la coherencia para la clase positiva



4.2. Tópicos para las K seleccionadas

4.2.1. NEGATIVOS

■ $K = 2$

TOPIC 0

flight, hour, time, customer, service, gate, seat, delayed, delay, would, customer_service, late, get, thanks, still, agent, amp, plane, back, hr.

TOPIC 1

flight, bag, airline, cancelled, get, plane, day, ua, problem, lost, please, way, need, worst, cancelled_flightled, flightled, luggage, want, amp, first.

■ $K = 6$

TOPIC 0

flight, hour, late, bag, plane, late_flight, cancelled, delay, hr, airline, delayed, worst, waiting, gate, ua, min, still, crew, ever, day.

TOPIC 1

service, customer, customer_service, get, flight, bag, need, time, tco,

airline, help, one, phone, please, thanks, amp, poor, still, make, lost.

TOPIC 2

flight, seat, cancelled, get, ua, delayed, cancelled_flightled, flightled, plane, refund, hour, one, connection, delay, ticket, weather, bag, would, day, getting.

TOPIC 3

flight, time, bag, gate, problem, check, last, agent, even, fly, booking, booking_problem, delayed, make, airline, get, seat, customer, minute, hour.

TOPIC 4

flight, tell, would, amp, tco, time, back, airline, airport, still, customer, day, response, like, dm, flying, going, status, week, take.

TOPIC 5

email, want, get, bag, number, trying, reservation, day, sent, supervisor, fail, someone, thing, claim, call, phone, like, could, still, problem.

■ **$K = 9$**

TOPIC 0

flight, cancelled, bag, get, delayed, hour, cancelled_flightled, flightled, day, amp, delay, due, ua, hr, issue, go, need, trying, plane, another.

TOPIC 1

flight, gate, delayed, seat, ua, plane, connection, ord, crew, amp, bag, hotel, time, ground, agent, min, gate_agent, late, miss, fail.

TOPIC 2

service, customer, customer_service, worst, ever, airline, experience, thanks, never, airport, baggage, time, terrible, agent, amp, flying, make, care, ewr, still.

TOPIC 3

still, get, bag, hour, waiting, flight, gate, would, dm, response, agent, plane, minute, tell, help, someone, back, phone, last, ua.

TOPIC 4

flight, late, hour, late_flight, problem, first, delay, tco, thanks, luggage,

flightr, late_flightr, plane, booking_problem, booking, weather, could, would, seat, one.

TOPIC 5

time, airline, ticket, get, tco, people, make, refund, yes, come, never, departure, would, last, better, bag, fly, flight, day, wait.

TOPIC 6

time, sure, help, know, leave, flight, lost, need, email, plane, one, get, customer, getting, response, number, someone, service, called, saying.

TOPIC 7

bag, airport, time, ua, asked, nothing, say, yet, ewr, hr, airline, got, lax, pilot, know, sfo, care, compensation, hung, staff.

TOPIC 8

airline, hour, like, plane, lt, please, bag, way, thing, flying, option, least, seat, tomorrow, look, without, flight, business, tco, delay.

4.2.2. POSITIVOS

■ $K = 3$

TOPIC 0

thanks, thank, flight, service, great, help, customer, much, good, today, time, gate, one, get, appreciate, tco, dm, ua, back, sent.

TOPIC 1

flight, thank, tco, amazing, made, airline, amp, thanks, pilot, problem, agent, love, day, year, new, nice, ua, way, customer, experience.

TOPIC 2

thanks, tco, thank, flight, got, bag, good, get, airline, care, seat, thnx, customer, happy, home, reply, still, lost, guy, issue.

■ $K = 7$

TOPIC 0

flight, thank, thanks, great, service, got, good, made, day, time, much, customer, tco, appreciate, home, ever, today, amp, best, airline.

TOPIC 1

flight, thanks, tco, made, amazing, thank, new, class, nice, gate, year, ua, crew, way, like, bag, customer, awesome, one, staff.

TOPIC 2

thank, thanks, help, sent, flight, customer, dm, ua, much, today, good, ok, bag, gate, know, get, service, reply, airline, appreciated.

TOPIC 3

thanks, flight, time, get, customer, good, help, care, great, service, awesome, bag, much, sure, last, yes, would, back, agent, staff.

TOPIC 4

tco, thank, thanks, flight, please, home, great, first, assistance, pilot, seat, service, make, amp, guy, got, wc, better, helping, agent.

TOPIC 5

thanks, flight, thnx, tco, customer, well, service, airline, amp, help, way, back, ever, get, still, best, ticket, train, one, send.

TOPIC 6

amazing, follow, thank, tco, one, great, please, dm, airline, back, airport, agent, flight, problem, thru, tampa, early, still, office, hour.

4.3. Elección del mejor numero de tópicos

4.3.1. Negativos

Para el numero de tópicos negativos hemos elegido el 9 ya que es el que mayor coherencia tiene y comparándolo con los demás es el que mejor separa los tópicos. Las razones que hemos sacado de este numero de tópicos son las siguientes:

cancelled flight, flight delayed, poor customer service, late flight, booking problem, lost bag

4.3.2. Positivos

De la misma manera en la que pasaba con nuestra compañía, en el caso de los comentarios positivos al coger un numero mas alto de tópicos estos empezaban a hablar de los mismos motivos. Por lo tanto, una vez mas, hemos decidido escoger 3 tópicos distintos.

Las razones serían: good customer service,help,gate agent / experience flight,pilot / bag care

5. Decisiones para crear la historia

5.1. Preproceso

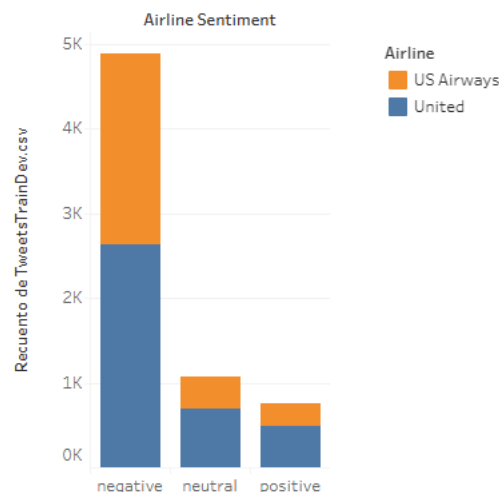
En el Tableau no hemos hecho un preproceso como tal, lo único que hemos hecho ha sido generar un csv con las instancias que pertenecen a nuestra empresa (US Airways) y a la empresa competente (United), para así facilitarnos a la hora de crear los gráficos, para no tener que generar un filtrado de nuestra empresa y la componte siempre que creemos un gráfico.

5.2. Generación de gráficos

A la hora de insertar los gráficos, como hay algunos que son muy grandes y no se pueden apreciar bien, hemos puesto un link en el titulo de cada gráfico, que te redirecciona a Tableau para poder visualizarlo mejor.

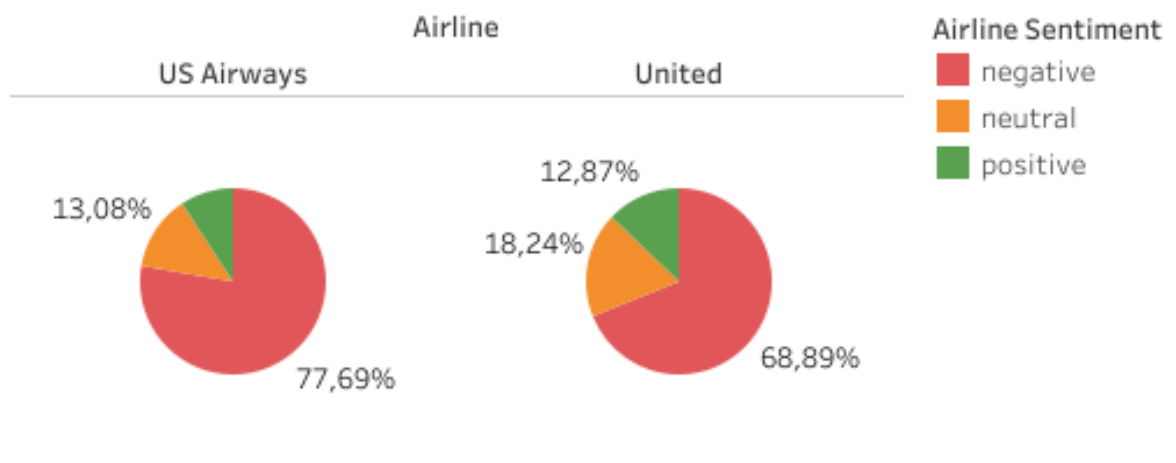
5.2.1. Gráfico 1

- **Descripción:** Utilizamos un gráfico de barras para analizar como están distribuidos los Tweets de nuestra compañía respecto a la competencia según como están clasificados por análisis de sentimiento.



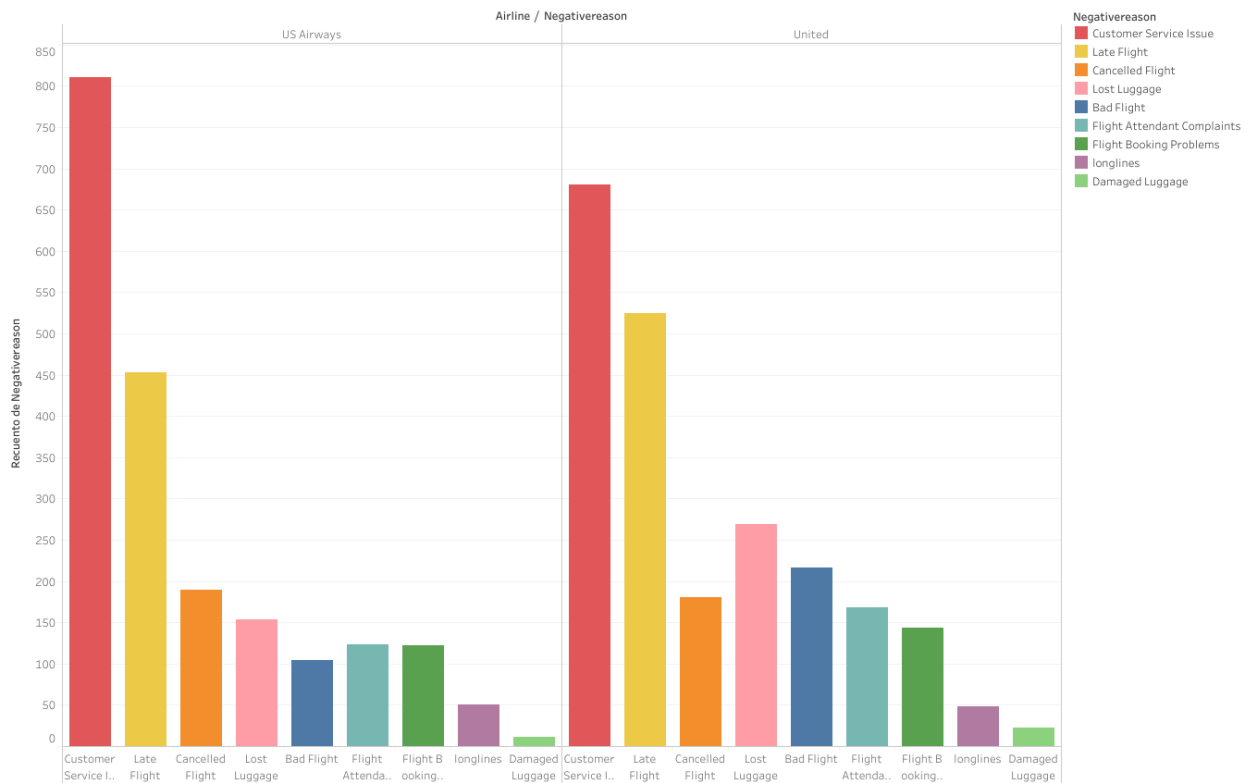
5.2.2. Gráfico 2

- **Descripción:** Este gráfico es igual que el anterior, solo que hemos utilizado un gráfico circular mostrando el porcentaje, para que se pueda ver mas fácil la diferencia de Tweets entre nuestra compañía y la competente.



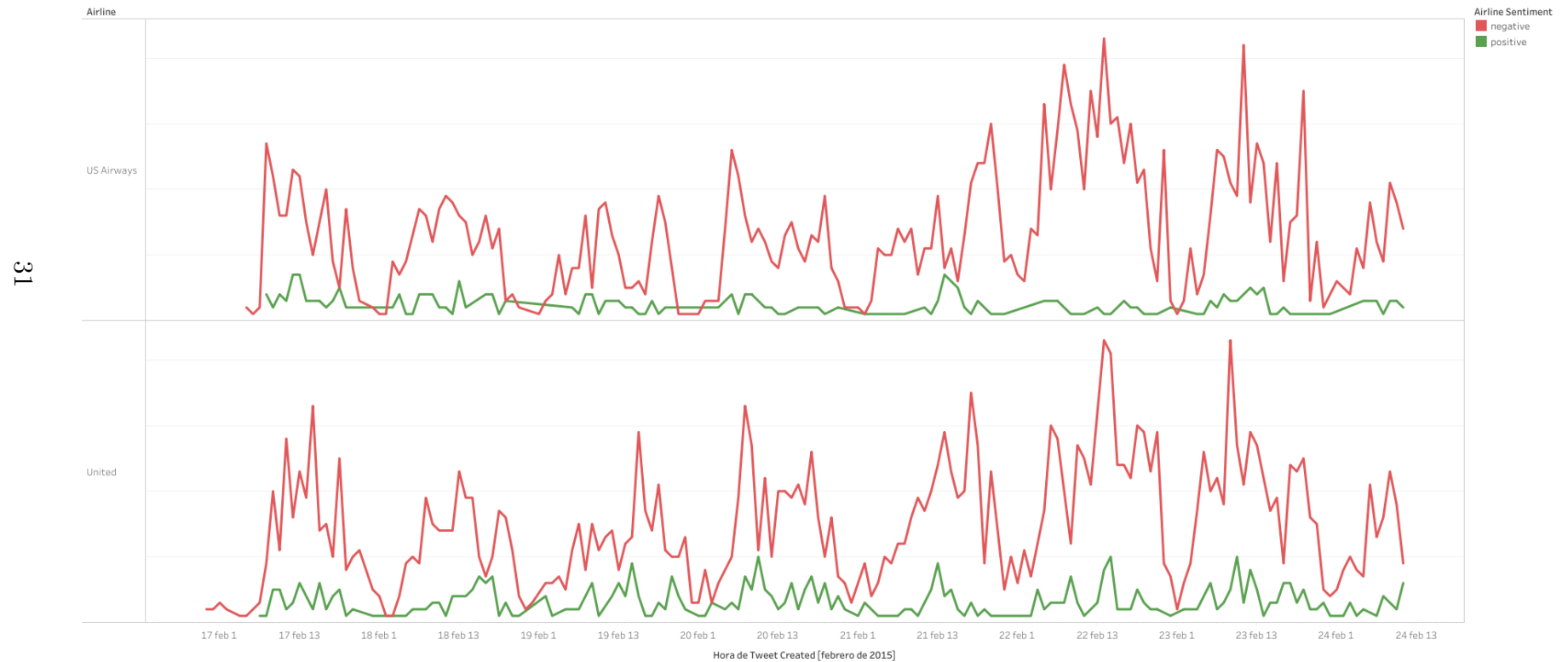
5.2.3. Gráfico 3

- **Descripción:** Utilizamos un gráfico de barras para analizar los distintos motivos que tienen los usuarios en cuanto a los Tweets negativos, de cara a conocer cuales son los aspectos que mas fallan, y por lo tanto, los que debemos mejorar con mayor importancia.



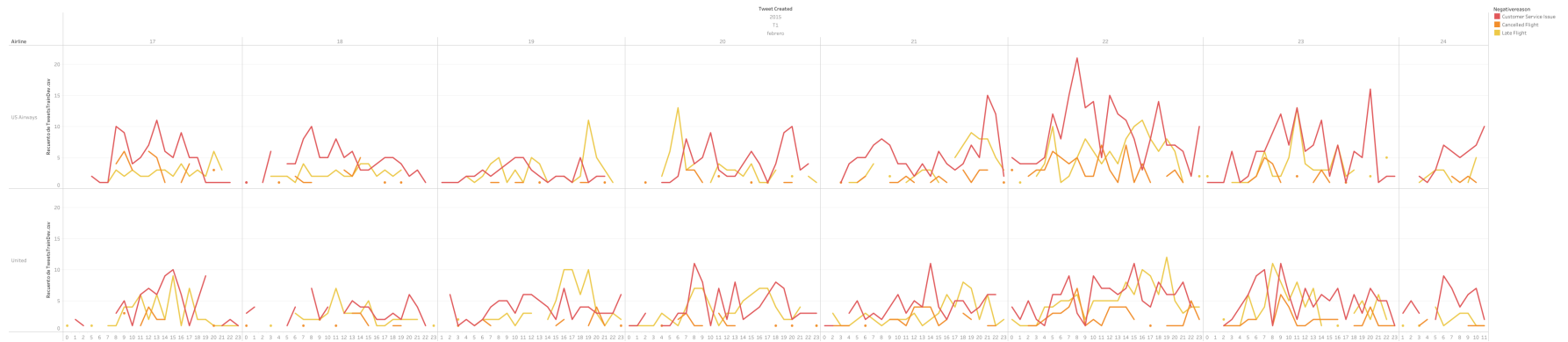
5.2.4. Gráfico 4

- **Descripción:** Utilizamos un gráfico lineal para comparar la cantidad de Tweets positivos y negativos de nuestra empresa respecto a la competente en el rango de días que se escribieron los Tweets.



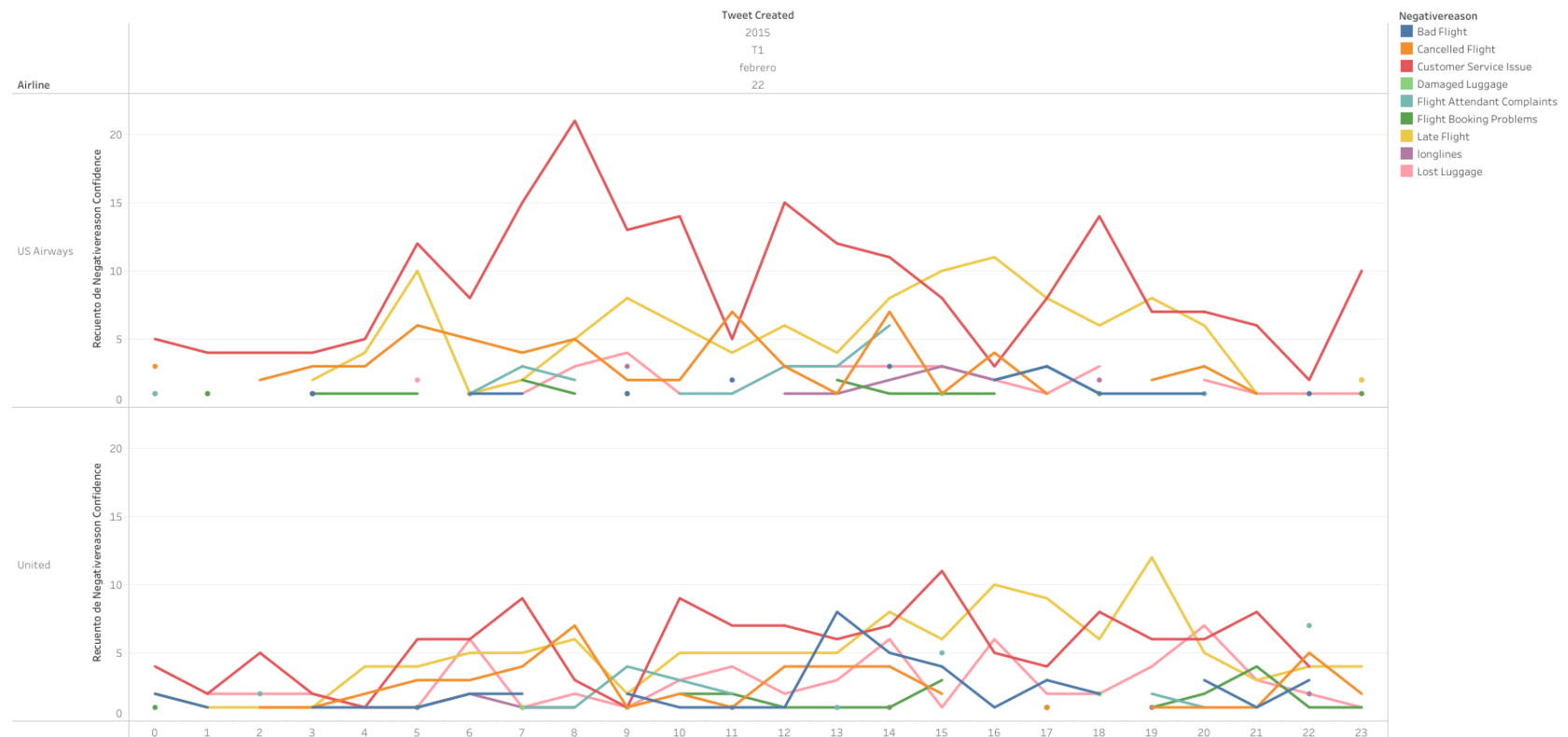
5.2.5. Gráfico 5

- **Descripción:** Utilizamos un gráfico lineal para comparar los distintos motivos que tienen los usuarios en cuanto a los Tweets negativos de nuestra empresa respecto a la competente en el rango de días que se escribieron los Tweets. Sabemos que hay mas razones negativas pero como había muchas y eso hacia que no se apreciase bien el gráfico y aparte eran menos significativas, hemos puesto las más significativas.



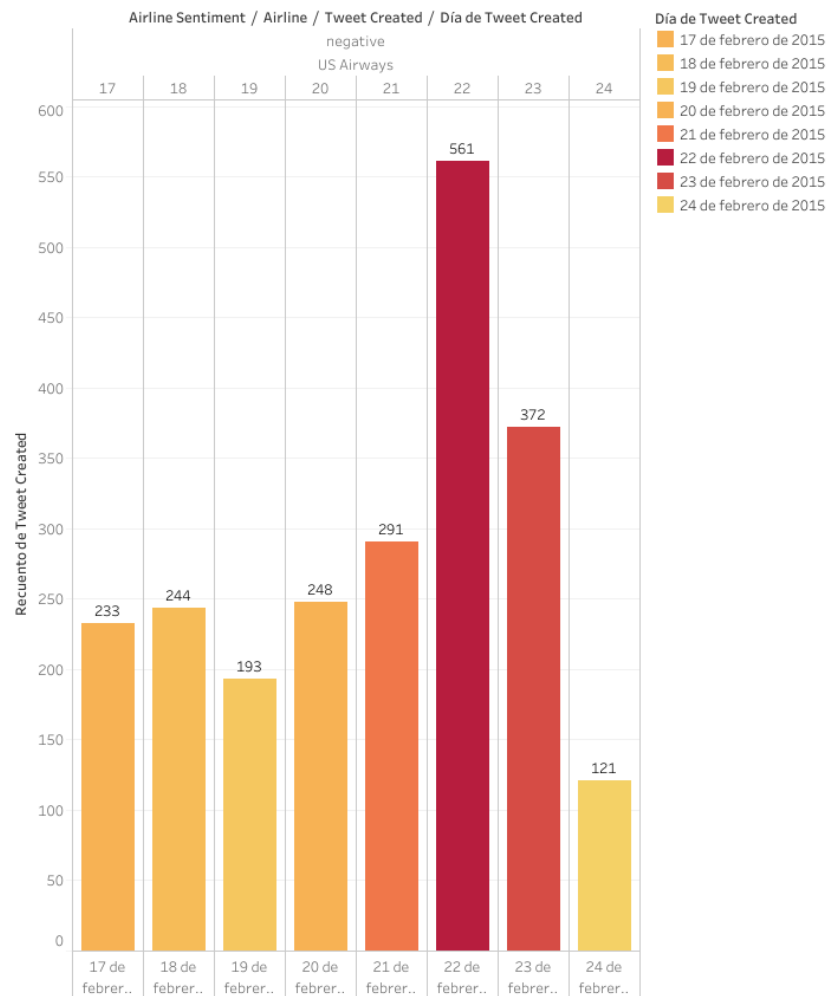
5.2.6. Gráfico 6

- **Descripción:** Utilizamos un gráfico lineal para comparar los distintos motivos que tienen los usuarios en cuanto a los Tweets negativos de nuestra empresa respecto a la competente el día 22 de febrero, ya que era el día que más Tweets se escribieron.



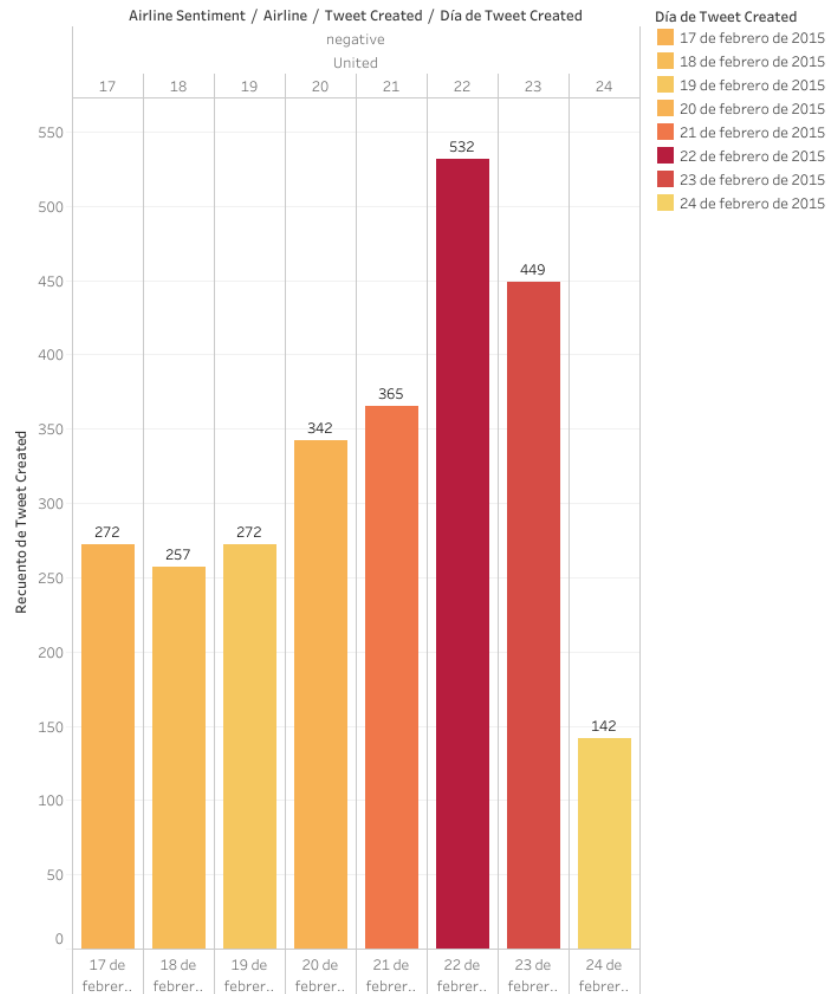
5.2.7. Gráfico 7

- **Descripción:** Utilizamos un diagrama de barras para analizar la cantidad de Tweets negativos que se han escrito en nuestra compañía US Airways en el rango de días que se escribieron los Tweets.



5.2.8. Gráfico 8

- **Descripción:** Utilizamos un diagrama de barras para analizar la cantidad de Tweets negativos que se han escrito en la compañía competente United en el rango de días que se escribieron los Tweets.



6. Conclusiones

Hemos sacado varias conclusiones después de realizar esta tarea. Nos hemos dado cuenta que las mismas razones por las que la mayoría se queja, hay algunos que están contentos. Esto se puede deber a que normalmente la gente tiende a tweetear las quejas y las criticas, antes que hacer una buena reseña. Otra de la razón puede ser que dependiendo del empleado que les haya tratado o el caso en particular, pueden variar las opiniones.

De todas formas, a pesar de que mucha gente da una buena reseña sobre un área, si hay otra mucha gente que se queja de lo mismo hay que intentar cambiar y mejoraren ese área. En el caso de nuestra empresa US Airways hemos comprobado que el servicio al cliente puede tener mucha mejora ya que es la razón por la que tenemos mas Tweets negativos. Destacamos los siguientes puntos a tratar con el jefe:

- Los domingos intentar tener mas empleados, hemos podido notar que hay un considerable incremento en el numero de criticas, sobre todo, hacia la atención al cliente que ofrece la empresa ese día y también por retrasos en los vuelos. A raíz de esto sospechamos que muy probablemente, esto se deba a una escasez de empleados anormal este día de la semana.
- Mejorar las llamadas telefónicas, no tener tanto tiempo en espera al cliente, esto puede deberse a que hay pocos empleados en ese área y no dan a basto. Hay que intentar bajar el tiempo que tardan hasta que te ponen con una persona real. También que no de corten las llamadas, cambiar de linea podría ser una solución.
- Otro objetivo a mejorar es la actitud de los empleados destinados a tratar con el cliente. Si bien hemos visto muchas buenas criticas hacia la actitud de estos empleados al cliente, hemos visto que muchas otras que son todo lo contrario. Esto seguramente se deba simplemente a que hay muchos empleados que cumplen su función correctamente y otros muchos que no.
- Intentar tener mejor calidad de las comidas del avión y ajustar los precios.

Por otro lado, si bien tenemos un ratio mayor de criticas negativas frente al rival, un punto positivo es que estas criticas están mayormente centradas en 2 tópicos, los cuales son servicio al cliente, y retrasos en los vuelos(Esto

es una ventaja frente a la competencia la cual tiene algo mas divididos los puntos en los que debe mejorar. Por lo que a priori podríamos destacar otras razones las cuales hace que la empresa vaya a peor, o al menos no centrarnos plenamente en ellas. En ellos estarían, vuelos cancelados o con retraso, mala gestión de las maletas y equipaje, y problemas al realizar reservas. Aquí se podrían tratar los siguientes puntos para intentar mejorar en estas áreas:

- Empleados mas profesionales y eficientes para evitar retrasos en los vuelos, principalmente en las puertas de embarque.
- Tener mejor control de las maletas y no tardar tanto en cargar y descargar las maletas. Se podría valorar la opción de tener mas empleados o mejor preparados para el tema del equipaje.
- Invertir mas en el correcto funcionamiento y diseño de la pagina web. Hemos podido ver que en el apartado que habla sobre los problemas de reservas y gestión de billetes de vuelo, una de las cosas sobre las que mas se habla es el mal funcionamiento de la pagina web, la cual tarda mucho en cargar y en avisar a los clientes sobre el estado tanto de su vuelo como su billete.

Como objetivo final de mejorar respecto a la competencia, tenemos sin duda que mejorar el área de atención al cliente, el cual es el que recibe en proporción muchísimas mas criticas que las de nuestro rival, tanto es así que a pesar de tener muchas menos criticas en prácticamente todas las demás áreas de la empresa nos pone casi con el mismo numero de criticas negativas generales que este a pesar de tener bastantes menos tweets en general.

Tenemos que destacar que nuestra experiencia de vuelo (Bad Flight) es mejor que la de la competencia (tenemos menos Tweets al respecto), esto seguramente se deba al tema de las azafatas, el piloto y la sensación dentro del avión. Creemos que es un punto fuerte que da valor a nuestra empresa y tenemos que seguir con la buena dinámica para estar por encima de United.