

Entrega Final Maestría en Economía

Federico Molina Magne

Viernes 14 de Diciembre

Índice

1. Introducción	2
2. Obtención y análisis de datos	2
2.1. El Gallito	3
2.1.1. Análisis temporal	3
2.1.2. Area de actividad y nivel técnico	6
2.1.3. Departamento	10
3. Pasos siguientes	10

1. Introducción

La presente entrega tiene por objetivo aproximarse a la caracterización de la demanda laboral por parte de las empresas en Uruguay en el periodo de noviembre a diciembre de 2018. Ello quiere decir, analizar cuales son los niveles técnicos solicitados, las áreas (dentro de la firma) y los departamentos a los cuales corresponden los avisos. Como adelanto, cerca del 90-95 % de las publicaciones corresponden al departamento de Montevideo, por lo cual, el análisis es mas bien a nivel departamental.

Dado que en Uruguay no existe ninguna encuesta referida a vacantes laborales solicitadas por las empresas, se opto por obtener los datos de avisos laborales publicados en internet, técnica conocida como ‘scraping’. Las fuentes de datos son dos, el portal web ‘el gallito’ y la página web computrabajo, mientras que la unidad de análisis es el aviso laboral. Pese a contar con dos fuentes de datos, en esta entrega solamente se muestran los datos obtenidos de el diario El País, sección de avisos laborales El Gallito, dada la extensión del trabajo se prefirió mostrar de forma extensa la información de dicha fuente. Pese a esto, el código de scraping de computrabajo se encuentra en el repositorio de github.

El presente trabajo, es el primer paso dentro de un proyecto superior cuyo objetivo es la creación de una serie de vacantes laborales, como proxy a la demanda laboral, representativa al menos de Montevideo. Para ello, deben ser identificados los avisos laborales que son repetidos tanto en el espacio como en el tiempo. Es decir, los avisos que se repiten en un intervalo de tiempo a determinar, los cuales serán considerados repetidos, y, los que se repiten en un mismo momento del tiempo en distintos sitios web pero que deben ser considerados una sola vez. Dicha tarea tiene la dificultad de que la estructura de las páginas web es distinta, por lo cual, el mismo aviso puede ser publicado de forma diferente. La forma más intuitiva de solucionarlo es identificar puesto, empresa, fecha y similitud, este último mediante análisis de texto (por ejemplo, similaridad de cosenos).

2. Obtención y análisis de datos

Respecto a las dificultades ya mencionadas en la introducción, aparece una adicional al realizar el scraping de cada página web. Existen dos opciones, la primera es iterar en cada portal web a través de cada página que contiene en general unos 20 avisos, entonces si hay 180 páginas completas existen 3600 avisos. La segunda, es obtener el link de cada aviso (iterando previamente por cada página) e iterar individualmente sobre cada link obteniendo toda la información detallada, sin embargo, aquí surge una dificultad adicional, ya que, la información detallada no sigue una estructura sino que se modifica en cada aviso laboral (esto es lo que sucede con computrabajo).

La estrategia utilizada fue primero iterar en cada portal web a nivel de página y obtener la información de cada aviso, incluido el link del aviso el cual actuará posteriormente como identificador. Luego realizar una exploración a nivel individual y guardar cada set de datos por separado. Esto último debido a que al momento de iterar por cada aviso individualmente, se produjeron errores en algunos links y el tiempo para llevar a cabo la tarea aumenta dramáticamente.

Los conjuntos de datos obtenidos del scraping son guardados a nivel local, dado que los avisos van desapareciendo de las páginas web a medida que pasa el tiempo, los mismos son guardados en un repositorio a nivel local. Los códigos utilizados están disponibles en el repositorio de github escrito anteriormente.

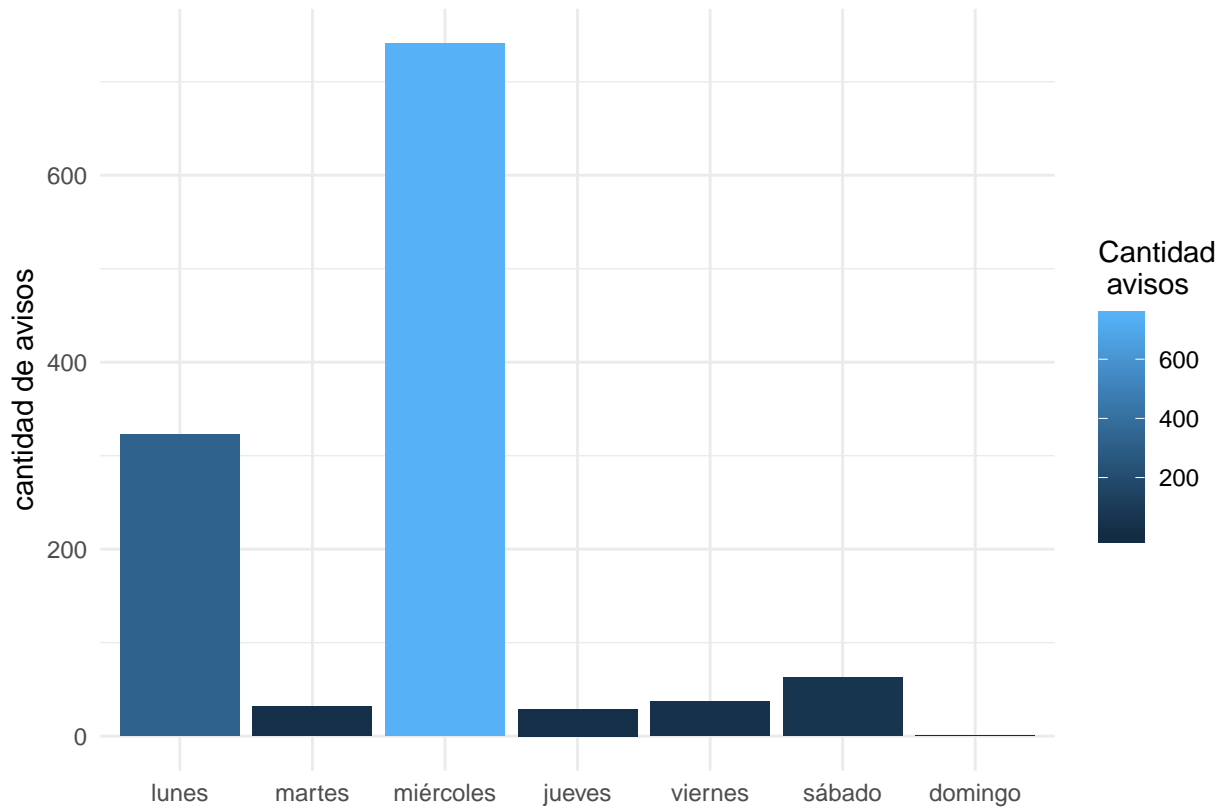


Figura 1: Avisos laborales por día de la semana

Se realiza una limpieza de datos de ambas páginas web el cual se encuentra en ver código limpieza de datos, dicha información es cargada posteriormente para realizar el análisis exploratorio de los datos. Los conjuntos de datos limpios son colgados en el repositorio de github de forma de poder recrear esta entrega.

2.1. El Gallito

Ver código scraping ‘El Gallito’.

El código referido a la obtención de los datos de la página web ‘el gallito’ pueden verse en el link superior. A continuación se analiza la información obtenida. Una observación a tener en cuenta es que si se trabaja con el archivo csv directamente deben transformarse algunas variables a factor. Si se usa el archivo rds ese problema desaparece.

2.1.1. Análisis temporal

El primer análisis realizado, ver figura 1, es en que días de la semana se concentra la publicación de avisos laborales en ‘El Gallito’. En este caso, los días con mayor cantidad de publicaciones son los miércoles y lunes, por el contrario el domingo es de menor cantidad de publicaciones. Es importante destacar que la fecha no es exacta dado que, a medida que transcurren los días en la página web aparece el texto “publicado hace 1 semana”, “publicado hace 2 semanas”, “publicado hace 1 mes”. Fueron dichos patrones los que se transformaron en segundos y los mismos fueron restados a la fecha de scraping obteniendo la supuesta fecha de publicación.

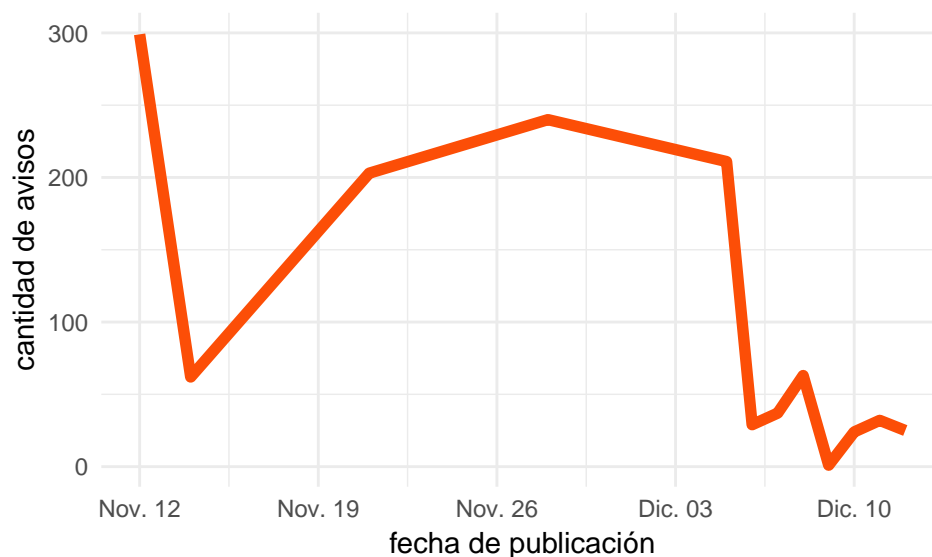


Figura 2: Avisos laborales por fecha de publicación

A continuación se crea una serie de tiempo desde noviembre a diciembre de los avisos publicados (figura 2). como se mencionó previamente, observar la fecha exacta de publicación no es posible (a menos que se realice una obtención de datos diaria o semanal), pero si puede realizarse una comparación a nivel semanal. Lo que se observa es que en los últimos días de noviembre se concentra la mayor cantidad de publicaciones. Es claro que no puede sacarse ninguna conclusión con solamente un mes de datos, en la medida que se sigan extrayendo datos se podrá ver si existe algún patron estacional a nivel mensual o semanal, dado que indicadores a nivel agregado de la actividad económica como el producto interno bruto o el empleo muestran una marcada estacionalidad, podría esperarse que las vacantes laborales también la tengan.

Dicha hipótesis se apoya en el análisis de datos de “El Gallito” entre 2013 y 2018, en cuya serie temporal parece existir cierta estacionalidad. Como se observa en la figura 3 hay picos aumentos y caídas marcados en todos los años, un claro indicio de estacionalidad, lo cual iría en línea con la teoría y los datos de variables relacionadas. Además, es importante notar la caída en la cantidad de avisos publicados, como se mencionó al comienzo el objetivo del trabajo (a mediano plazo) es obtener una serie de vacantes laborales como un proxy a la demanda laboral por parte de las empresas. Por esto, podría pensarse que si la serie de El Gallito es medianamente representativa, la demanda laboral debería haber disminuido notoriamente. Sin embargo, la representatividad de El Gallito en cuanto a publicaciones laborales ha caído de forma importante a lo que fuera 10 o 20 años atrás. Actualmente páginas como LinkedIn, computrabajo y buscojobs logran captar igual o mayor cantidad de publicaciones.

Dado que parece existir estacionalidad, la serie de vacantes fue descompuesta de forma aditiva en 3 componentes mediante la función decompose: tendencia, estacionalidad e irregular. Como se aprecia en la figura 4 la serie presenta una marcada estacionalidad y una tendencia a la baja muy importante. Los picos a la baja en el componente estacional se dan principalmente en el último mes del año, durante todos los años. Es decir, diciembre suele ser el periodo con un efecto estacional fuertemente negativo, esto es mismo es lo que se observa incipientemente en los datos obtenidos por scraping, pero dado que los datos obtenidos son hasta el 12 de diciembre no se logra captar completamente el fenómeno.

Es importante destacar que la información de la figura 3 es de caracter confidencial, ha sido facilitada por el diario El País la información de todos los avisos laborales publicados en “El Gallito” entre 2013 y 2018, para uso con fines puramente académicos y no se ha autorizado su publicación.

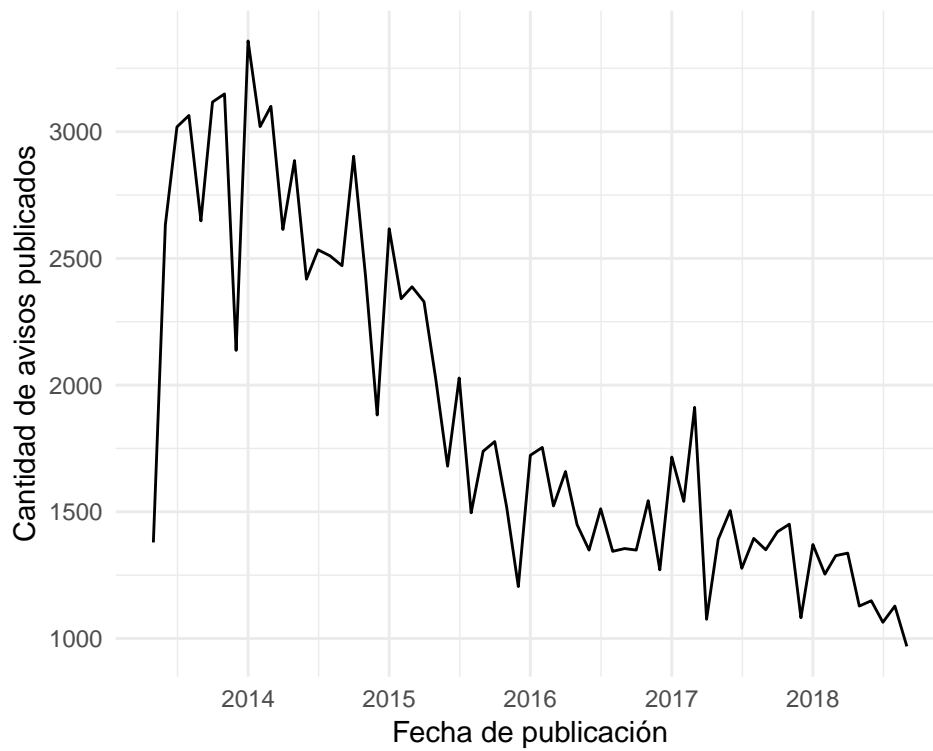


Figura 3: Serie temporal vacantes laborales entre 2013-2018 El Gallito

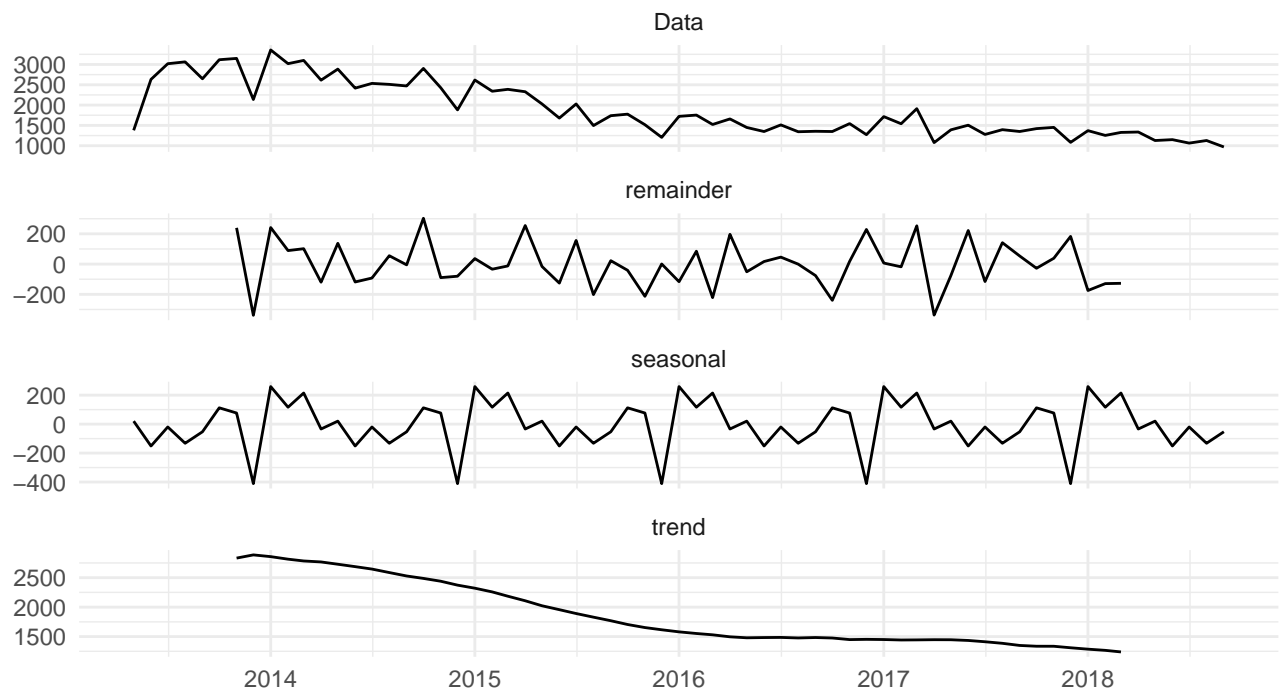


Figura 4: descomposición de series de vacantes en componente de tendencia, estacionalidad, estacional e irregular.

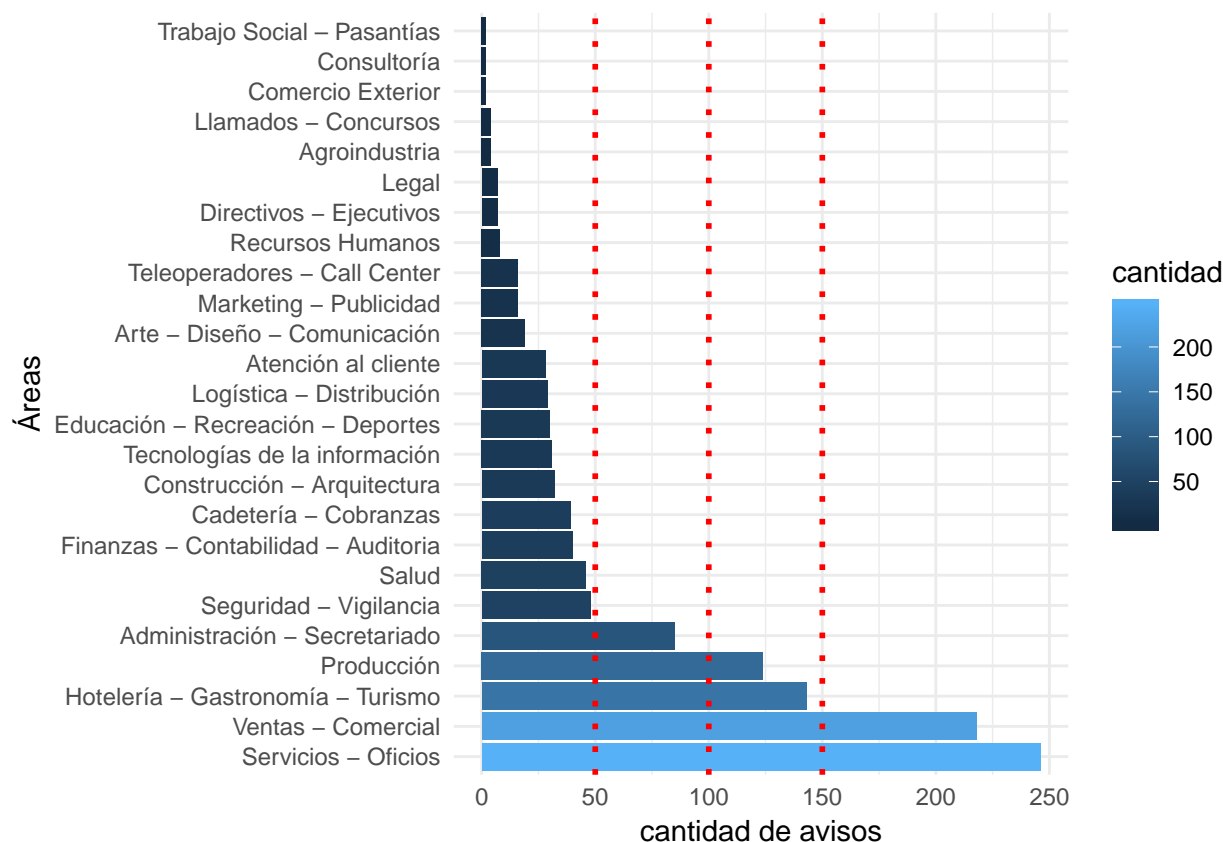


Figura 5: Áreas de actividad solicitadas por las empresas

Lo que se ha hecho es trabajar dichos datos y obtener una serie temporal de la cantidad total de avisos publicados en el periodo 2013-2018 sin excluir avisos duplicados. Es esta información, la referida solamente a la serie temporal, la que ha sido compartida en el repositorio de github en formato rds con fines de reproducción.

Por último, no han sido depurado los avisos laborales que se repiten en el tiempo, dado que para ello es necesario definir un intervalo de tiempo sobre el cual buscar. Una opción sería seguir al índice HWOL de los EEUU y utilizar un periodo de tiempo mensual. Pese a esto, la dinámica de la serie no debería verse alterada de forma relevante.

2.1.2. Area de actividad y nivel técnico

Es de interés que sector de actividad concentra cada página web, o si la representación de los mismos tiende a ser uniforme. En el caso de “El Gallito” existe un claro sesgo hacia sectores de baja calificación como ventas y servicios no financieros. Entre ‘Ventas - Comercial’ y ‘Servicios - Oficios’ se concentra cerca del 38 % de los avisos.

Si bien, se observa claramente este patrón en la figura 5, sería interesante saber si el mismo es producto de que otras páginas web han logrado captar las publicaciones de puestos de mayor calificación. Una hipótesis es que sitios como LinkedIn apuntan a un mercado de mayor preparación, de hecho, es razonable dado que buscojobs nació como un portal web para captar puestos laborales del sector tecnológico recibiendo financiación por parte de la Camara Uruguaya de Tecnología e Información (CUTI).

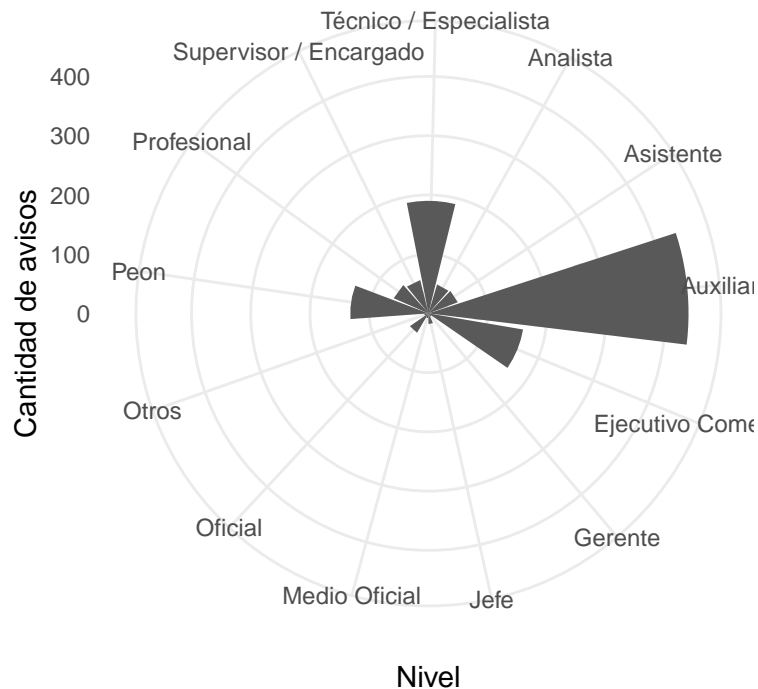


Figura 6: Nivel técnico solicitado por las empresas

Para reforzar la idea anterior de que las publicaciones de “El Gallito” tienen un sesgo hacia sectores y trabajadores con baja calificación, la figura 6 es extremadamente ilustrativo. En el mismo se observan los distintos niveles técnicos (la definición de dichos niveles corresponde a el diario El País) solicitados por las empresas, desde gerentes hasta peones y auxiliares. Claramente el nivel más solicitado es el de auxiliar seguido por ejecutivo comercial, técnico - especialista y peón.

En base al cuadro 1 los llamados de auxiliares representan cerca del 36 % de los avisos totales, siendo por lejos los niveles técnicos más solicitados, seguido por técnico - especialista con un 15 %, ejecutivo comercial con un 13 % y peón con un 11 %. Por otro lado jefe es solicitado en un 1 % de los avisos y gerente no llega a representar un 1 %.

Vale aclarar que puede haber al menos dos efectos, primero el ya mencionado sesgo de baja calificación, pero por otro lado es esperable que los puestos de menor calificación tengan mayor cantidad de avisos, ya que, en una empresa por lo general abunda el personal de menor calificación en contra posición a gerentes o jefes. Obviamente esto es menos notorio en algunos sectores como software y tecnología, pero incluso en ellos el patrón se repite.

Sería interesante comparar con los llamados publicados en linkedIn y ver si los porcentajes muestran valores similares o difieren fuertemente.

Para finalizar el análisis exploratorio de área de actividad y nivel técnico, se visualizan ambos de forma conjunta. De donde surge que los puestos solicitados en el área de ventas y comercial son mayormente ejecutivos comerciales y el de menor cantidad es jefes. Dicho análisis conjunto puede observarse en la figura 7.

Como se observa en el cuadro 2 los porcentajes correspondientes a cada sector y area de actividad suman uno por fila, es decir, por área de actividad económica solicitada por las empresas. Por lo tanto, la tabla se debe leer de la siguiente manera: del total de avisos publicados pertenecientes a ventas - comercial el

Cuadro 1: Porcentaje por nivel técnico

Nivel	Freq
Gerente	0.00
Otros	0.01
Medio Oficial	0.01
Jefe	0.01
Oficial	0.03
Analista	0.04
Asistente	0.04
Supervisor / Encargado	0.05
Profesional	0.05
Peon	0.11
Ejecutivo Comercial	0.13
Técnico / Especialista	0.15
Auxiliar	0.36

Cuadro 2: Porcentaje por área de actividad desagregado por nivel solicitado

	Analista	Asistente	Auxiliar	Ejecutivo Comercial	Gerente	Jefe	Medio Oficial	Oficial	Otros	Peon	Profesional	Supervisor / Encargado	Técnico / Especialista
Administración - Secretariado	0.04	0.20	0.74	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Agroindustria	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.75	0.00	0.00
Arte - Diseño - Comunicación	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.05	0.00	0.05
Atención al cliente	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cadetería - Cobranzas	0.00	0.10	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Comercio Exterior	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50
Construcción - Arquitectura	0.00	0.06	0.00	0.00	0.00	0.12	0.00	0.09	0.00	0.03	0.22	0.12	0.34
Consultoría	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Directivos - Ejecutivos	0.00	0.29	0.00	0.00	0.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Educación - Recreación - Deportes	0.00	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.67	0.03	0.20
Finanzas - Contabilidad - Auditoría	0.08	0.40	0.32	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.02	0.00
Hotelería - Gastronomía - Turismo	0.00	0.00	0.54	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.41
Legal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Llamados - Concursos	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Logística - Distribución	0.03	0.00	0.72	0.00	0.00	0.03	0.00	0.00	0.00	0.17	0.00	0.03	0.00
Marketing - Publicidad	0.19	0.12	0.31	0.00	0.06	0.12	0.00	0.00	0.06	0.00	0.00	0.00	0.12
Producción	0.00	0.02	0.10	0.00	0.00	0.03	0.05	0.16	0.00	0.23	0.10	0.11	0.20
Recursos Humanos	0.25	0.00	0.12	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.12	0.12
Salud	0.00	0.02	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.30
Seguridad - Vigilancia	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Servicios - Oficinas	0.00	0.00	0.27	0.00	0.00	0.00	0.01	0.06	0.00	0.39	0.00	0.00	0.26
Tecnologías de la información	0.74	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.10
Teleoperadores - Call Center	0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00
Trabajo Social - Pasantías	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ventas - Comercial	0.00	0.00	0.14	0.73	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.12	0.00

73 % de los mismos fueron pidiendo puesto para un nivel de ejecutivos comercial, un 14 % correspondió a auxiliar, un 12 % a supervisor - encargado y un 1 % a jefe. Del total de avisos correspondientes a servicios - oficios el 39 % fueron peon, el 26 % técnico - especialista, el 27 % auxiliar, un 6 % oficial y un 1 % medio - oficial.

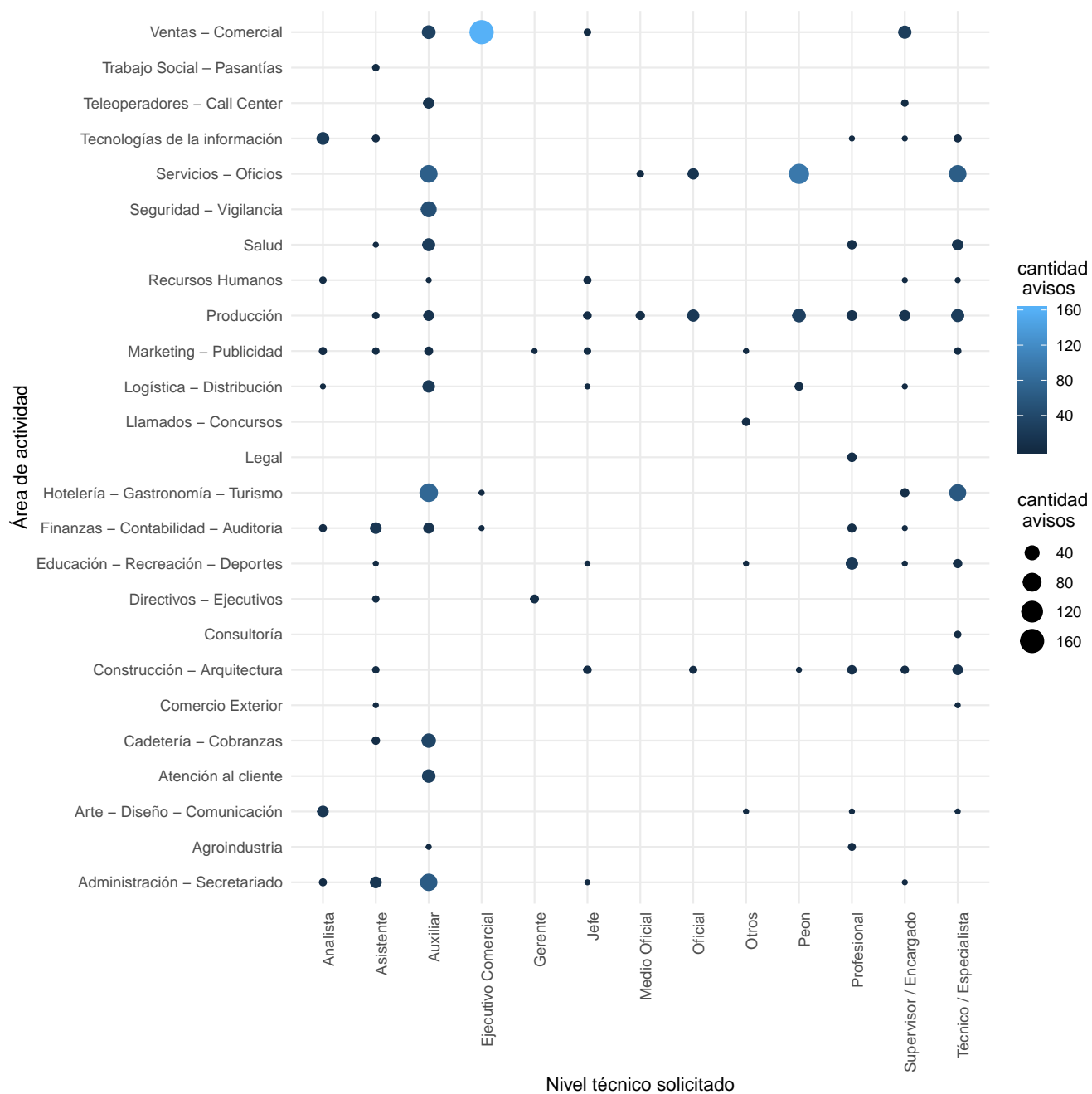


Figura 7: Nivel técnico y actividad solicitado por las firmas

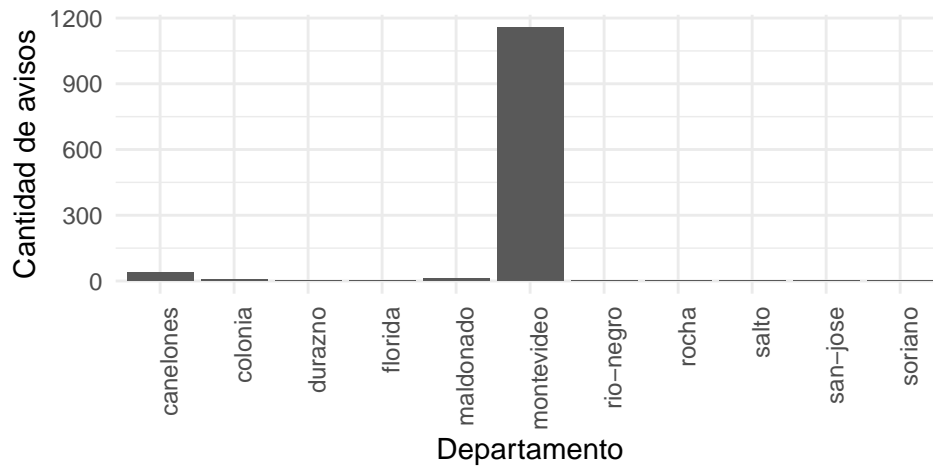


Figura 8: Departamentos al que pertenecen las firmas que solicitan los puestos laborales

2.1.3. Departamento

Finalmente es importante conocer a que departamento pertenecen o donde están localizadas las empresas que publican avisos laborales en el gallito. Según los datos tomados entre noviembre y diciembre de 2018, el 94 % son de Montevideo, un 3 % son de Canelones y el resto corresponde a los restantes departamentos, como se observa en la figura 8. Por lo tanto, la información obtenida de “El Gallito” solo debería ser usada para aproximarse a conclusiones respecto a la demanda laboral de Montevideo. Es de remarcar que esto es solo una aproximación, ya que, actualmente dicho portal web no debe considerarse representativo ni siquiera a nivel departamental, esto debido a la existencia de diferentes páginas web que publican avisos laborales tales como computrabajo o linkedIn.

3. Pasos siguientes

A partir de este análisis exploratorio de los datos obtenidos mediante scraping de El Gallito, y de su análisis entre 2013-5 y 2018-9 se observa una caída permanente en la cantidad de publicaciones, una fuerte estacionalidad con un pico fuertemente negativo en diciembre. Dada la caída tendencial, puede pensarse que si las vacantes laborales son un proxy de la demanda laboral, esta última estaría cayendo de forma sistemática. Sin embargo, si bien dicha hipótesis no puede descartarse a priori, la misma también puede ser explicada por la menor representatividad por parte del diario El País en lo que respecta al mercado de avisos laborales en contra posición a un aumento de nuevos competidores como LinkedIn, computrabajo y buscojobs, entre otros.

De los 1226 avisos obtenidos, el 95 % pertenece a Montevideo, por lo cual, los datos solo pueden utilizarse para realizar inferencia para dicho departamento.

Queda seguir con este EDA tanto para El Gallito como Computrabajo. Seguir obteniendo datos para poder ampliar el periodo de análisis, a la vez que terminar de obtener la información individual de cada aviso, procesar la misma y terminar armando una tabla normalizada para portal. Finalmente los datos de las dos páginas deben ser combinados sin que se repitan avisos laborales. Para ello es necesario tener además de la identificación por puesto, empresa y fecha de publicación, un indicador de similaridad entre avisos construido mediante text mining.