



Universidade do Minho
Escola de Ciências

Relatório TP1 e TP2

AMBIENTES E LINGUAGENS DE PROGRAMAÇÃO PARA CIÊNCIA DE DADOS

Licenciatura em Ciência de Dados

Docentes: Tiago Batista & Nuno Macedo

GRUPO 2

GABRIELA SOARES - A112282

LARA SACRAMENTO - A110333

RODRIGO ALVES - A111408

Índice

- [1. Introdução](#)
- [2. Bibliotecas e Tecnologias Utilizadas](#)
- [3. Descrição Geral da Aplicação](#)
- [4. Funcionalidades](#)
 - [4.1. Listagem e Pesquisa de Ofertas de Emprego \(TP1- A e B\)](#)
 - [4.2. Análise Textual e Competências Técnicas \(TP1- C e D / TP2-C\)](#)
 - [4.3. Enriquecimento de Ofertas e Análise por Empresa \(TP2- A\)](#)
 - [4.4. Estatísticas e Análise Agregada por Zona Geográfica \(TP2- B\)](#)
 - [4.5. Exportação de Dados e Persistência de Resultados \(TP2- D\)](#)
- [5. Tratamento de Erros e Validações](#)
- [6. Conclusão](#)

1. Introdução

No âmbito da unidade curricular **Ambientes e Linguagens de Programação para Ciência de Dados (ALPCD)**, foi desenvolvido um projeto prático com o objetivo de aplicar conceitos de programação em Python, consumo de serviços web, processamento de dados e desenvolvimento de aplicações de linha de comandos.

O projeto encontra-se organizado em dois momentos complementares, designados **Trabalho Prático 1 (TP1)** e **Trabalho Prático 2 (TP2)**, que culminam no desenvolvimento de uma aplicação única, integrada e progressivamente enriquecida.

No **TP1**, o foco incidiu sobre a interação com a **API pública do itjobs.pt**, permitindo a recolha e análise de ofertas de emprego na área das tecnologias de informação. Foram implementadas funcionalidades para a listagem de anúncios, pesquisa por empresa e localização, inferência do regime de trabalho e contagem de competências técnicas (*skills*), bem como a exportação dos resultados para ficheiros CSV. Estas funcionalidades estabeleceram a base estrutural e funcional da aplicação.

O **TP2** surge como uma evolução direta do trabalho anteriormente desenvolvido, reutilizando a mesma arquitetura e expandindo as suas capacidades. Neste segundo momento, a aplicação foi enriquecida através da integração com a plataforma **Teamlyzer**, recorrendo a técnicas de *web scraping* para obter informação adicional sobre empresas e perfis profissionais, nomeadamente classificações, benefícios e competências técnicas mais relevantes. Adicionalmente, foram introduzidas funcionalidades de análise estatística agregada e exportação avançada de dados.

Desta forma, o projeto final corresponde a uma **ferramenta de linha de comandos modular e extensível**, que combina dados provenientes de diferentes fontes, permitindo uma análise mais completa e contextualizada do mercado de trabalho na área das tecnologias de informação, cumprindo os objetivos definidos nos enunciados do TP1 e do TP2.

2. Bibliotecas e Tecnologias Utilizadas

No desenvolvimento da aplicação foram utilizadas diversas bibliotecas da linguagem **Python**, selecionadas de acordo com os requisitos do projeto e com os objetivos definidos nos enunciados do TP1 e do TP2. A escolha destas bibliotecas permitiu garantir uma implementação modular, legível e adequada ao processamento de dados e à interação com serviços externos.

A biblioteca **Typer** foi utilizada para a construção da interface de linha de comandos (CLI). Esta biblioteca possibilita a definição clara de comandos, argumentos e opções, bem como a geração automática de ajuda e validação básica de parâmetros, contribuindo para uma utilização intuitiva da aplicação.

Para a comunicação com serviços externos via HTTP, nomeadamente com a **API do itjobs.pt** e com a plataforma **Teamlyzer**, foi utilizada a biblioteca **requests**. Esta biblioteca permitiu a realização de pedidos HTTP do tipo GET, o tratamento das respostas obtidas e a gestão de erros associados a falhas de comunicação ou códigos de resposta inválidos.

A biblioteca **json** foi utilizada para o processamento e apresentação dos dados devolvidos pela API do [itjobs.pt](#) e pelas funcionalidades de enriquecimento, permitindo a leitura, manipulação e impressão de informação em formato estruturado. Complementarmente, a biblioteca **csv** foi utilizada para a exportação dos resultados para ficheiros CSV, facilitando a persistência e análise posterior dos dados.

No âmbito da análise textual, foram utilizadas as bibliotecas **re** e **unicodedata**. A biblioteca **re** permitiu a aplicação de expressões regulares para a identificação de padrões relevantes, como regimes de trabalho e competências técnicas, enquanto a biblioteca **unicodedata** foi utilizada para a normalização de texto, assegurando comparações mais robustas entre cadeias de caracteres.

A biblioteca **datetime** foi utilizada para a validação e comparação de datas, nomeadamente na análise de competências técnicas dentro de intervalos temporais definidos pelo utilizador. Adicionalmente, a biblioteca **operator** foi utilizada para facilitar a ordenação de resultados com base em critérios específicos.

Para o *web scraping* da plataforma **Teamlyzer**, foi utilizada a biblioteca **BeautifulSoup** (**bs4**), que permitiu analisar a estrutura HTML das páginas, localizar elementos relevantes e extrair a informação necessária de forma estruturada.

Por fim, foram utilizadas bibliotecas auxiliares da biblioteca padrão do Python, como **sys** e **urllib.parse**, para o controlo da execução da aplicação e para a codificação adequada de parâmetros em URLs.

A combinação destas bibliotecas permitiu desenvolver uma aplicação robusta e extensível, capaz de integrar múltiplas fontes de dados e de responder de forma eficaz aos requisitos definidos nos dois trabalhos práticos.

3. Descrição Geral do Programa

O **programa desenvolvido** consiste numa **interface de linha de comandos (CLI)** implementada em **Python**, que permite consultar, analisar e enriquecer dados relacionados com ofertas de emprego na área das tecnologias de informação. O programa foi concebido de forma modular, recorrendo à biblioteca **Typer**, o que possibilita a definição clara de comandos independentes e promove uma utilização intuitiva por parte do utilizador.

Como fonte principal de dados, o programa utiliza a **API pública do itjobs.pt**, através da qual são obtidas informações detalhadas sobre anúncios de emprego, incluindo título da oferta, empresa, localização, descrição, tipo de contrato e data de publicação. Estes dados constituem a base para as funcionalidades de listagem, pesquisa e análise implementadas inicialmente no **TP1**.

No **TP2**, o programa foi estendido de modo a integrar uma segunda fonte de informação, a plataforma **Teamlyzer**, recorrendo a técnicas de *web scraping*. Esta integração permite

enriquecer os dados obtidos a partir do itjobs.pt com informação adicional sobre empresas e perfis profissionais, nomeadamente classificações, benefícios, descrições institucionais e competências técnicas mais relevantes associadas a determinados cargos.

O programa encontra-se organizado por **comandos**, cada um associado a uma funcionalidade específica. Os resultados são apresentados, por defeito, em formato **JSON**, facilitando a leitura estruturada da informação. Sempre que aplicável, é igualmente disponibilizada a opção de **exportação para ficheiros CSV**, permitindo a posterior análise, reutilização ou integração dos dados recolhidos noutras ferramentas.

Esta abordagem garante uma separação clara de responsabilidades, facilita a manutenção do código e possibilita a expansão futura do programa, mantendo a coerência entre as funcionalidades desenvolvidas no TP1 e as extensões introduzidas no TP2.

4. Funcionalidades

4.1. Listagem e Pesquisa de Ofertas de Emprego (TP1- A e B)

Uma das funcionalidades base da aplicação consiste na **listagem e pesquisa de ofertas de emprego**, implementada no âmbito do TP1 e posteriormente reutilizada como suporte para funcionalidades mais avançadas no TP2.

Na funcionalidade de **listagem de ofertas**, é possível obter os **N anúncios mais recentes** publicados no itjobs.pt, recorrendo à respetiva **API REST**. O utilizador define o número de ofertas a listar, sendo efetuada **validação do argumento fornecido**, garantindo que o valor indicado é positivo. Os resultados são apresentados **no terminal em formato JSON**, correspondendo diretamente aos dados devolvidos pela API, incluindo campos como o título da oferta, empresa, localização, data de publicação e descrição.

Adicionalmente, foi implementada uma funcionalidade de **pesquisa de ofertas part-time**, permitindo filtrar anúncios com base em **três critérios principais**: localidade, empresa e tipo de contrato. Para esse efeito, a aplicação analisa os tipos de contrato associados a cada oferta, recorrendo a **expressões regulares** para identificar anúncios classificados como *part-time*. O número máximo de resultados a apresentar é igualmente controlado pelo utilizador.

Em ambas as funcionalidades, é disponibilizada a opção de **exportação dos resultados para ficheiros CSV**, contendo os campos definidos no enunciado do TP1 (título, empresa, descrição, data de publicação, salário e localização). Esta funcionalidade permite a análise posterior dos dados recolhidos fora do contexto da aplicação de linha de comandos, reforçando a utilidade prática da solução desenvolvida.

Estas funcionalidades constituem a **base da aplicação**, assegurando a recolha estruturada de dados a partir do itjobs.pt, os quais são posteriormente reutilizados e enriquecidos nas funcionalidades introduzidas no TP2.

```

PS C:\Users\rodri\OneDrive - Universidade do Minho\Ambiente de Trabalho\git_reps\ALPCD_4\TP1> python jobscli.py top 3
[{"id": 507927, "company": {"id": 984, "name": "KCS IT", "logo": "https://static.itjobs.pt/images/companies/d9/3ed/984/logo.png", "description": "A KCS IT \xe9 uma empresa de consultoria, outsourcing e formac\u00e3o, especializada em servi\u00e7os de gest\u00e3o de projectos profissionalizada alinhados com as metodologias e melhores pr\u00e1ticas do PMI.", "address": "Rua Sousa Martins, n\u00famero 10 \r\n1050-218 Lisboa - Portugal", "phone": "+351 213 174 164", "url": "http://www.kcsit.pt", "url_twitter": "https://twitter.com/KCS_IT", "url_facebook": "https://www.facebook.com/pages/kcs-it/168449752382", "url_linkedin": "https://www.linkedin.com/company/kcsit", "slug": "kcs-it"}, {"companyId": 984, "title": "Java Developer", "body": "<p><strong>We're looking for the special, unique and amazing YOU!</strong></p><p>@ KCS IT, we look for the ones that stands out, for those that always wants to be better and fight for it, and for those who has the same values that we do: dedication, energy, integrity, transparency, flexibility, trust, honesty, hard work, proactivity, team work.</p><p>At KCS we stand for equality and value diversity. We create a safe, diverse environment where opportunities are equal for everyone! We do not discriminate based on age, ethnicity, sexual orientation, gender, disability or any factor other than merit. All applications with skills for the position are welcome!</p><p>We are looking for an Amazing:</strong> Java Developer</p><p><strong>Seniority: </strong>Mid (3-5 years)</p><p><strong>Type of position:</strong> Hybrid in Lisbon. The candidates must be in Portugal.</p><p><strong>The amazing you, will have:</strong></p><ul><li>4+ years of experience as a Fullstack Developer.</li><li>Strong proficiency with Java (v.+8), including Spring or Spring Boot frameworks.</li><li>Hands-on experience with cloud technologies (AWS, Azure, or GCP).</li><li>Good understanding of microservices architecture and REST APIs.</li><li>Experience with relational and/or NoSQL databases (e.g., PostgreSQL, MySQL, MongoDB).</li><li>Familiarity with CI/CD tools (Jenkins, GitLab CI, GitHub Actions, etc.).</li><li>Strong problem-solving skills and attention to detail.</li><li>Fluent in English (spoken and written).</li><li><strong>Nice to Have:</strong> Knowledge of containerization (Docker, Kubernetes).</li><li>Experience with reactive programming (e.g., WebFlux).</li><li>Previous work in Agile/Scrum environments.</li><li>Familiarity with testing frameworks for both frontend and backend.</li></ul><strong>Why should you become part of our family?</strong><ul><li><u>You can develop a career that fits you</u>: your career development is personalized, taking in consideration your needs and goals from a short to long term</li><li><u>Interesting Challenges Ahead</u>: you can work for several clients from different sectors of activity</li><li><u>Free training programs</u>: Our training and certification programs in languages, tech, behavior and business will help you to reach your full potential faster</li><li><u>International projects in Benelux</u>: you can gain international experience working in different countries</li></ul>"}]
```

Figura 1 - Execu\u00e7\u00e3o do comando top

4.2. An\u00e1lise Textual e Compet\u00eancias T\u00e9cnicas (TP1- C e D / TP2-C)

Para al\u00e9m da simples listagem de ofertas, a aplicac\u00e3o desenvolvida inclui funcionalidades de **an\u00e1lise textual**, cujo objetivo \u00e9 extrair informa\u00e7\u00e3o relevante a partir das descri\u00e7\u00e3es dos an\u00fcncios e de fontes externas. Estas funcionalidades foram introduzidas no TP1 e posteriormente aprofundadas no TP2, mantendo uma l\u00f3gica comum de an\u00e1lise de compet\u00eancias t\u00e9cnicas (*skills*).

No **TP1**, foram implementadas duas funcionalidades centrais neste dom\u00ednio. A primeira permite a **infer\u00eancia do regime de trabalho** (remoto, h\u00edbrido, presencial ou outro) a partir de um identificador de oferta (*job_id*). Esta infer\u00eancia baseia-se na an\u00e1lise do texto do t\u00fatulo e da descri\u00e7\u00e3o do an\u00fcncio, recorrendo a express\u00e3es regulares para detetar padr\u00f5es associados a cada regime. Sempre que a an\u00e1lise textual n\u00f3o \u00e9 conclusiva, s\u00f3o utilizados campos adicionais disponibilizados pela API do itjobs.pt, como a indica\u00e7\u00e3o de trabalho remoto e a exist\u00eancia de localiza\u00e7\u00e3es associadas \u00e0 oferta.

A segunda funcionalidade do TP1 relacionada com an\u00e1lise textual consiste na **contagem de ocorr\u00eancias de compet\u00eancias t\u00e9cnicas (skills)** nas descri\u00e7\u00e3es das ofertas de emprego, dentro de um intervalo temporal definido pelo utilizador. Para tal, as descri\u00e7\u00e3es dos an\u00fcncios s\u00f3o analisadas atrav\u00e9s de express\u00e3es regulares, permitindo identificar e quantificar a presen\u00e7a de um conjunto pr\u00e9-definido de tecnologias e linguagens de programac\u00e3o. Os resultados s\u00f3o apresentados em formato JSON, ordenados por n\u00famero de ocorr\u00eancias.

No **TP2**, esta abordagem é estendida e aprofundada através da integração com a plataforma **Teamlyzer**, permitindo uma análise de competências técnicas baseada não apenas nas descrições dos anúncios, mas também em dados agregados recolhidos externamente. A funcionalidade correspondente permite obter as **principais competências técnicas associadas a um determinado perfil profissional**, devolvendo as *top N skills* mais relevantes, bem como o respetivo número de ocorrências.

Para este efeito, a aplicação recorre a técnicas de *web scraping*, identificando o perfil profissional mais adequado (*profession role*) no Teamlyzer e extraíndo as tecnologias associadas a esse perfil. Esta abordagem complementa a análise realizada no TP1, oferecendo uma perspetiva mais estruturada e global das competências mais procuradas para cada função.

Desta forma, as funcionalidades de análise de texto e competências técnicas evoluem de uma abordagem local, baseada exclusivamente nas descrições dos anúncios do itjobs.pt (TP1), para uma abordagem enriquecida e agregada, suportada por dados externos do Teamlyzer (TP2). Esta progressão evidencia a coerência entre os dois trabalhos práticos e demonstra a capacidade de integrar múltiplas fontes de dados para uma análise mais completa do mercado de trabalho.

```
PS C:\Users\laras\Desktop\ALPCD_4\TP1> python jobscli.py type 507886  
remoto
```

Figura 2- Execução comando type

```
PS C:\Users\rodri\OneDrive - Universidade do Minho\Ambiente de Trabalho\git repos\ALPCD_4\TP1> python jobscli.py list skills "Data Scientist"  
[  
  {  
    "skill": "python",  
    "count": 51  
  },  
  {  
    "skill": "sql",  
    "count": 25  
  },  
  {  
    "skill": "amazon-web-services-aws",  
    "count": 20  
  },  
  {  
    "skill": "machine-learning",  
    "count": 17  
  },  
  {  
    "skill": "apache-spark",  
    "count": 16  
  },  
  {  
    "skill": "azure",  
    "count": 14  
  },  
  {  
    "skill": "artificial-intelligence",  
    "count": 13  
  },  
  {  
    "skill": "scala",  
    "count": 11  
  },  
  {  
    "skill": "java",  
    "count": 9  
  },  
  {  
    "skill": "cloud",  
    "count": 8  
  }]
```

Figura 3- Execução comando list skills

4.3. Enriquecimento de Ofertas e Análise por Empresa (TP2- A)

Após a implementação das funcionalidades base de recolha e análise de ofertas no TP1, o TP2 introduz uma extensão significativa da aplicação através do **enriquecimento dos dados de cada oferta de emprego**, recorrendo a informação externa proveniente da plataforma **Teamlyzer**.

No TP1, a análise das ofertas de emprego encontra-se limitada à informação disponibilizada pela **API do itjobs.pt**, sendo possível obter dados como o título da oferta, empresa, descrição, localização, tipo de contrato e data de publicação. Embora esta informação seja suficiente para análises básicas, não fornece contexto adicional sobre a empresa empregadora, como a sua reputação, benefícios oferecidos ou níveis salariais médios.

Com o **TP2 (alínea a)**, esta limitação é ultrapassada através da implementação de um mecanismo de **integração entre o itjobs.pt e o Teamlyzer**. A partir de um identificador de oferta (*job_id*), a aplicação obtém os dados base da oferta através da API do itjobs.pt e, de seguida, procura a página correspondente da empresa no Teamlyzer. Sempre que possível, esta procura é efetuada de forma direta utilizando o *slug* da empresa; caso contrário, é utilizado um método de pesquisa baseado no ranking do Teamlyzer.

Uma vez identificada a empresa no Teamlyzer, a aplicação recorre a técnicas de *web scraping* para extrair informação adicional relevante, nomeadamente:

- classificação média da empresa (*rating*);
- descrição textual da empresa, extraída a partir de metadados da página;
- benefícios associados;
- referências a intervalos salariais praticados, quando disponíveis.

Estes dados são então **integrados no objeto JSON da oferta original**, resultando num output enriquecido que combina informação proveniente de múltiplas fontes. Esta funcionalidade representa uma evolução direta das operações realizadas no TP1, mantendo o mesmo padrão de utilização (consulta por *job_id*), mas acrescentando uma camada adicional de contexto e valor analítico.

Adicionalmente, tal como nas funcionalidades do TP1, é disponibilizada a opção de **exportação dos resultados para ficheiro CSV**, permitindo a persistência da informação enriquecida para análise posterior. Esta funcionalidade contribui para a coerência global da aplicação e reforça a reutilização dos mecanismos de exportação já implementados.

```

PS C:\Users\rodri\OneDrive - Universidade do Minho\Ambiente de Trabalho\git repos\ALPCD_4\TP1> python jobscli.py get 507927
{
    "id": 507927,
    "company": {
        "id": 984,
        "name": "KCS IT",
        "logo": "https://static.itjobs.pt/images/companies/d9/3ed/984/logo.png",
        "description": "A KCS IT é uma empresa de consultoria, outsourcing e formação, especializada em serviços de gestão de projectos profissionalizados alinhados com as metodologias e melhores práticas do PMI.",
        "address": "Rua Sousa Martins, nº10 \r\n1050-218 Lisboa - Portugal",
        "phone": "+351 213 174 164",
        "url": "http://www.kcsit.pt",
        "url_twitter": "https://twitter.com/KCS_IT",
        "url_facebook": "https://www.facebook.com/pages/kcs-it/168449752382",
        "url_linkedin": "https://www.linkedin.com/company/kcsit",
        "slug": "kcs-it"
    },
    "companyId": 984,
    "title": "Java Developer",
    "body": "<p><strong>We're looking for the special, unique and amazing YOU!</strong></p><p>@ KCS IT, we look for the ones that stands out, f or those that always wants to be better and fight for it, and for those who has the same values that we do: dedication, energy, integrity, transparency, flexibility, trust, honesty, hard work, proactivity, team work.</p><p>At KCS we stand for equality and value diversity. We create a safe, diverse environment where opportunities are equal for everyone! We do not discriminate based on age, ethnicity, sexual orientation, gender, disability or any factor other than merit. All applications with skills for the position are welcome!</p><p><strong>We are looking for an Amazing:</strong> Java Developer</p><p><strong>Seniority: </strong>Mid (3-5 years)</p><p><strong>Type of position:</strong> Hybrid in Lisbon. The candidates must be in Portugal.</p><p><strong>The amazing you, will have:</strong></p><ul><li>4+ years of experience as a Fullstack Developer.</li><li>Strong proficiency with Java (v.+8), including Spring or Spring Boot frameworks.</li><li>Hands-on experience with cloud technologies (AWS, Azure, or GCP).</li><li>Good understanding of microservices architecture and REST APIs.</li><li>Familiarity with CI/CD tools (Jenkins, GitLab CI, GitHub Actions, etc.).</li><li>Strong problem-solving skills and attention to detail.</li><li>Fluent in English (spoken and written).</li><li>Nice to Have:</strong> Knowledge of containerization (Docker, Kubernetes).</li><li>Experience with reactive programming (e.g., WebFlux).</li><li>Previous work in Agile/Scrum environments.</li><li>Familiarity with testing frameworks for both frontend and backend.</li></ul><p><strong>Why should you become part of our family?</strong><ul><li><u>You can develop a career that fits you:</u> your career development is personalized, taking in consideration your needs and goals from a short to long term</li><li><u>Interesting Challenges Ahead:</u> you can work for several clients from different sectors of activity</li><li><u>Free training programs:</u> Our training and certification programs in languages, tech, behavior and business will help you to reach your full potential faster</li><li><u>International projects in Benelux:</u> you can gain international experience in Benelux and balance a new way of living with work</li><li><u>Type of projects </u>(depending on the project you might find one of this types of projects):<ul><li><u>In hybrid Systems:</u> Is important to balance work with socialization, that's why a hybrid system works for us and for you</li><li><u>Full Remote projects:</u> If you want to work while you enjoy the comfort of your home</li><li><u>Full Onsite projects:</u> if you prefer the company of your colleagues!</li></ul><li><u>Take care of your well-being:</u> Enjoy our free nutrition, psychology, general medicine appointments and our yoga and personal training days... all remote</li></ul><p><strong>Who are we? </strong></p><p>Founded in 2008 and based in Lisbon, KCS IT is a consulting company in the field of Information Technology and Services, focused on creating

```

Figura 4 - Execução comando get

4.4. Estatísticas e Análise Agregada por Zona Geográfica (TP2- B)

Para além da análise individual de ofertas de emprego, a aplicação desenvolvida permite também realizar **análises agregadas**, com o objetivo de identificar padrões globais no mercado de trabalho. Esta funcionalidade corresponde à **alínea b) do TP2** e assenta diretamente nos mecanismos de recolha de dados implementados no **TP1**.

No TP1, as funcionalidades de listagem e pesquisa permitem obter conjuntos alargados de ofertas de emprego a partir da API do itjobs.pt. Estes dados constituem a base necessária para a realização de análises estatísticas mais abrangentes, embora no TP1 os resultados sejam apresentados essencialmente ao nível individual de cada anúncio.

Com o **TP2 (alínea b)**, esta abordagem é estendida através da implementação de uma funcionalidade que permite **agregar e contabilizar vagas por zona geográfica e por tipo de trabalho**. Para tal, a aplicação recolhe um conjunto de ofertas de emprego através da API do itjobs.pt e procede à análise dos campos de localização associados a cada anúncio. As ofertas são então agrupadas por região (zona) e por tipo ou nome da posição, sendo calculado o número total de vagas para cada combinação.

Os resultados desta análise são apresentados em formato **JSON**, permitindo uma visualização clara das regiões com maior número de oportunidades e dos tipos de trabalho mais frequentes em cada zona. Adicionalmente, é disponibilizada a opção de **exportação para ficheiro CSV**, contendo as colunas definidas no enunciado do TP2: zona, tipo de trabalho e número de vagas.

Esta funcionalidade representa uma evolução natural das operações realizadas no TP1, uma vez que reutiliza os dados obtidos através da API do itjobs.pt, mas aplica técnicas de **agregação e síntese de informação**, permitindo uma perspetiva mais global e analítica do mercado de trabalho.

```
PS C:\Users\rodri\OneDrive - Universidade do Minho\Ambiente de Trabalho\git_reps\ALPCD_4\TP1> python jobscli.py statistics Braga
[{"zona": "Braga", "tipo_trabalho": "Cybersecurity Engineer", "vagas": 1}, {"zona": "Braga", "tipo_trabalho": "Java Backend Developer Senior", "vagas": 1}, {"zona": "Braga", "tipo_trabalho": "Programador HTML/CSS/JS (.NET & SQL Server)", "vagas": 1}, {"zona": "Braga", "tipo_trabalho": "Cloud Engineer", "vagas": 1}, {"zona": "Braga", "tipo_trabalho": "Frontend Developer - Angular", "vagas": 1}, {"zona": "Braga", "tipo_trabalho": "Junior Information Security Consultant/Auditor", "vagas": 1}, {"zona": "Braga", "tipo_trabalho": "Fullstack Developer - Java & Angular", "vagas": 1}, {"zona": "Braga", "tipo_trabalho": "Backend Developer - Java", "vagas": 1}]
```

Figura 5 - Execução do comando statistics

4.5. Exportação de Dados e Persistência de Resultados (TP2- D)

Ao longo do desenvolvimento da aplicação, foi dada especial atenção à possibilidade de **persistir os resultados obtidos**, permitindo a sua reutilização e análise fora do contexto da execução da CLI. Esta funcionalidade de exportação encontra-se presente desde o TP1 e é posteriormente estendida e generalizada no TP2, correspondendo à **alínea d)** do segundo trabalho prático.

No **TP1**, a exportação para ficheiros **CSV** é disponibilizada nas funcionalidades de listagem e pesquisa de ofertas de emprego. Sempre que o utilizador o solicita, os resultados obtidos a partir da API do itjobs.pt são guardados em ficheiros CSV contendo campos como título da oferta, empresa, descrição, data de publicação, salário e localização. Esta funcionalidade permite uma análise posterior dos dados, por exemplo em folhas de cálculo ou outras ferramentas de análise de dados.

No **TP2**, este mecanismo de exportação é reutilizado e alargado a novas funcionalidades. Em particular, a exportação para CSV é aplicada:

- ao **enriquecimento de ofertas com dados do Teamlyzer** (TP2 a), permitindo guardar informação adicional sobre a empresa, como classificações, benefícios e referências salariais;
- às **estatísticas agregadas por zona e tipo de trabalho** (TP2 b), gerando ficheiros CSV com colunas específicas definidas no enunciado;
- à **análise de competências técnicas (skills)** obtidas a partir do Teamlyzer (TP2 c), possibilitando a persistência das *top skills* associadas a um determinado perfil profissional.

Desta forma, a exportação para CSV assume um papel transversal em toda a aplicação, funcionando como um mecanismo comum que suporta tanto as funcionalidades base do TP1 como as extensões introduzidas no TP2. Esta abordagem contribui para a coerência global do projeto e reforça a sua utilidade prática.

5. Tratamento de Erros e Validações

De modo a garantir a robustez e fiabilidade da aplicação desenvolvida, foram implementados diversos mecanismos de **validação de dados e tratamento de erros**, tanto no TP1 como no TP2. Estes mecanismos visam assegurar que a aplicação reage de forma controlada a entradas inválidas, falhas de comunicação com serviços externos e situações excepcionais durante a execução.

No que respeita às **validações de entrada**, a aplicação verifica, por exemplo, se o número de ofertas a listar é maior que zero, se os identificadores de ofertas (*job_id*) são fornecidos corretamente e se os intervalos de datas utilizados na análise de competências técnicas seguem o formato esperado e respeitam a ordem cronológica. Sempre que estas condições não são cumpridas, é apresentada uma mensagem de erro clara ao utilizador, evitando a execução incorreta das funcionalidades.

Relativamente à **comunicação com a API do itjobs.pt**, são verificados os códigos de resposta devolvidos pelo servidor, sendo igualmente tratadas situações em que a API retorna mensagens de erro explícitas. Desta forma, evita-se o processamento de respostas inválidas e assegura-se que o utilizador é informado de eventuais problemas na obtenção dos dados.

No **TP2**, devido à integração com o Teamlyzer através de técnicas de *web scraping*, foi necessário reforçar o tratamento de exceções associadas a pedidos HTTP, como falhas de ligação, *timeouts* ou alterações inesperadas na estrutura das páginas. Sempre que a informação pretendida não pode ser obtida, a aplicação lida com essas situações de forma controlada, devolvendo valores nulos ou mensagens informativas, sem comprometer a execução global do programa.

De um modo geral, estes mecanismos de validação e tratamento de erros contribuem para a estabilidade da aplicação e para uma melhor experiência de utilização, sendo

consistentes com as boas práticas de desenvolvimento de software abordadas na unidade curricular.

6. Conclusão

Com o desenvolvimento deste projeto foi possível aplicar, de forma integrada, os conceitos abordados na unidade curricular **Ambientes e Linguagens de Programação para Ciência de Dados**, resultando numa aplicação de linha de comandos funcional, modular e extensível.

No **Trabalho Prático 1**, foram estabelecidas as bases da aplicação, através da implementação de mecanismos de consumo da **API do itjobs.pt**, permitindo a listagem, pesquisa e análise de ofertas de emprego, bem como a inferência de informação relevante a partir de texto e a exportação de dados para ficheiros CSV. Estas funcionalidades permitiram explorar técnicas fundamentais de programação em Python, manipulação de dados em formato JSON e utilização de expressões regulares.

O **Trabalho Prático 2** constituiu uma evolução direta do trabalho anteriormente desenvolvido, acrescentando uma camada de **enriquecimento e análise avançada de dados**. A integração com a plataforma **Teamlyzer**, recorrendo a técnicas de *web scraping*, permitiu complementar a informação proveniente do itjobs.pt com dados adicionais sobre empresas e perfis profissionais, bem como realizar análises agregadas e aprofundadas, nomeadamente ao nível de estatísticas por zona geográfica e competências técnicas mais relevantes.

A junção das funcionalidades do TP1 e do TP2 resultou numa aplicação coerente e consistente, capaz de combinar dados provenientes de múltiplas fontes e de apresentar resultados estruturados em formatos adequados à análise posterior. O projeto evidencia a importância da reutilização de código, da organização modular e da integração de diferentes tecnologias na construção de soluções práticas para análise de dados.

Como trabalho futuro, seria possível melhorar a robustez dos mecanismos de *web scraping*, explorar técnicas adicionais de análise de texto e aumentar a escalabilidade da aplicação, nomeadamente através da paginação de resultados e da integração com outras fontes de dados relevantes para o estudo do mercado de trabalho.