IOAKEIM HADJIMPALASIS

https://ioakeim-h.github.io/
ioakeim.h@gmail.com
357 - 99250369

PROFILE

Ioakeim is a self-taught programmer with professional experience from reputable organisations in the UK and Cyprus. He excels in advanced data analytics - extracting insights and leveraging machine learning algorithms. His automation proficiency and commitment to best coding practices deliver efficient, user-friendly software solutions.

EXPERIENCE

IT Consultant - Digital Innovation Team, KPMG (CY) Feb 2022 - Present

Automating Dataset Cleaning Operations for an Insurance Company

- Restructured and modularized code, automating tasks for efficiency.
- Streamlined code flow, ensuring a user-friendly experience.
- Enhanced code maintainability through clear logic articulation.
- Successfully delivered optimized and maintainable code to the client.

Data Preparation and Predictive Analytics for a Communication Service Provider

- Debugging and testing substantial volumes of existing Python code used for pre-processing, model fitting and learning.
- Created a script that leveraged historical data to generate new features, calculating product churn and product switch recursively every three months over a four-year period.
- Developed a rudimentary fbprophet model to assess its potential for forecasting revenue.
- Employed Pandas to construct contingency tables for analyzing product competition among the client and other vendors.
- Evaluated data quality by producing correlation plots using Matplotlib.

Recalculation for Financial Reporting in the Telco and Forex Sectors

- Comparing system populations with SAP Accounting using Microsoft SQL Server - importing data from local files, creating tables and performing aggregations.
- Performing data cleaning using regular expressions, creating new features, and conducting table joins with Pandas, while effectively communicating data issues to the client and seeking feedback on the implementation of logic.

Finalization of a Segmentation Tool for an Energy and Petroleum Company

- Managed the ETL pipeline by adjusting the extraction connection from the SQL database and incorporated relative paths for flexible CSV file exports.
- Resolved code errors and converted the tool into an executable for enhanced usability.
- Maintained constant communication with the client about project challenges and solutions.

Data Wrangling for the Digital Transformation of a Financial Services Company

- Conducted basic data checks on R-generated data using Microsoft SQL Server.
- Performed data manipulation tasks such as table joins from CSV files using Pandas.

Professional Placement - Leadership Team, ONS (UK) Jun 2021 - Sep 2021

- Used R to analyse Census data from 2011 and 2021. Decided on suitable variables, cleaned these and produced descriptives that helped senior leadership diagnose potential issues.
- Performed a literature review about the mental health of veterans in England and provided a concise summary of the main findings using tables and references.

PROJECTS

DataGlitch: A Python Toolkit for Messy Data

DataGlitch is a Python package that addresses data challenges in pandas DataFrames. It efficiently detects mixed data types using regular expressions and extracts subsets based on numeric, ambiguous, and non-numeric values. DataGlitch offers seamless handling of non-ASCII characters, providing options for their replacement or translation. With its powerful fuzzy string matching capabilities, DataGlitch enables effective data exploration by facilitating column and value searches within datasets.

IOAKEIM HADJIMPALASIS

https://ioakeim-h.github.io/
ioakeim.h@gmail.com
357 - 99250369

PROJECTS

Miniseries: Data Engineering for Machine Learning with Python (GitHub)

Missing Data

- Using regular expressions and visualization techniques to explore missingness.
- Extracting the LittleMCAR function from R with rpy2 to determine the missing mechanism.
- Multiple imputation of missing categorical data using deep learning (DataWig).

Categorical Data

- Quantifying and reducing cardinality.
- Popular and evidence-based encoding techniques and potential pitfalls.

TECHNICAL SKILLS

Programming: An experienced Python programmer working with Pandas and Numpy on a daily basis. Proficiency in other programming languages such as R. Good knowledge of HTML and CSS.

Machine Learning: Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, k-means, PCA, Association Rule Learning, Causal Impact Analysis.

Relational Databases: Integrating SQL with Python to perform CRUD operations and extract data for analysis. Basic knowledge of theoretical concepts such as normalisation and denormalisation.

CORE SKILLS

Communication: Collaborating with colleagues and clients on a daily basis to resolve complex technical issues and develop plans of action. My diverse background has given me a unique perspective on problem-solving, especially in understanding the needs and motivations of individuals. I believe this skillset is highly advantageous in a technical role, allowing me to communicate effectively with both technical and non-technical stakeholders.

Teamwork: I'm adaptable, open to feedback, and prioritize producing clear and efficient work. My creativity and collaboration skills have led to successful outcomes on every project. I value and incorporate the ideas and perspectives of my team.

Problem-solving: As a self-taught programmer, I possess a proven track record of effectively extracting valuable insights from cluttered information. I excel at developing actionable plans to address poorly defined problems, and remain patient and adaptable in the face of unexpected events. My decision-making is always grounded in both evidence and experience, allowing me to consistently deliver optimal results.

EDUCATION

MSc Psychology, Cardiff University — 2020-2021

- Developed a thesis that focused on associative learning in humans. Data were analysed using two one-way Mixed ANOVAs and several Pearson Correlations. Results were compared against previous literature and were discussed in detail.
- Provided live sessions for master's students who requested support with statistics. Explained theoretical and practical topics and shaped the learning experience according to individual needs and capabilities.

BSc Sports and Exercise Science, University of Essex — 2017-2020

- Designed the methodological procedure of a study, recruited and tested participants in a laboratory, analysed the data collected, discussed findings and successfully developed a scientific report.

COURSES & CERTS

CS50's Introduction to Programming with Python, edX

Syllabus: Functions, Variables | Conditionals | Loops | Exceptions | Libraries | Unit Tests | File I/O | Regular Expressions | Object-Oriented Programming | Et Cetera

DATA SCIENCE INFINITY, Andrew Jones

Syllabus: SQL | Python (Numpy, Pandas, Matplotlib) | AB Testing | Machine Learning (Data Preparation, Supervised & Unsupervised Models) | Advanced Applications of Scikit-learn